Electronic Companion

3 Years, 2 Papers, 1 Course Off: Optimal Non-Monetary Reward Policies

Authors: Shivam Gupta, Wei Chen, Milind Dawande, Ganesh Janakiraman

Remark: Throughout this electronic companion, we adopt the notation that, for a random variable X and an event A, $\mathbb{E}[X;A]=\mathbb{E}[X|A]\cdot\mathbb{P}[A]$.

Appendix A: Proofs of Results in Section 4

Proof of Lemma 1: Fix an arbitrary limited-term reward policy with parameters (t_1,t_2,m). First, we obtain an upper bound on the expected long-run per-period utility of the agent for an *arbitrary* effort policy in response to the given reward policy (Lemma A.1). We develop this upper bound for the case where the output of the agent is a stochastic function of his effort. Thus, the same bound also holds for the special case where the agent's output is deterministic, given his effort. Second, for the case where the agent's output is deterministic, we identify an effort policy which achieves this bound (Lemma A.2).

LEMMA A.1. *Assume that the output of the agent is stochastic – that is, given an input effort $\lambda_t\in\mathbb{R}_+$ by the agent in period t, his output is $q_t=w_t\cdot\lambda_t$, where w_t is a realization of a random variable $W_t\in\mathbb{R}_+$. Under the limited-term reward policy (t_1,t_2,m), the expected long-run per-period utility of the agent is upper bounded as follows: (i) If $r\geq m/(t_1+t_2-1)$, then the upper bound is $v-c\cdot m/(t_1+t_2-1)$. (ii) If $r<m/(t_1+t_2-1)$, then the upper bound is 0.*

Proof of Lemma A.1. Under the limited-term reward policy (t_1,t_2,m), for any period t and any effort policy $\lambda\in\Lambda$, we have:

$$
\begin{aligned}
&\mathbb{E}\left[\pi_t^{LT}\left(W_1\lambda_1,W_2\lambda_2,\ldots,W_{t-1}\lambda_{t-1}\right)\right]\\
&=\mathbb{E}\left[\pi_t^{LT}\left(W_1\lambda_1,W_2\lambda_2,\ldots,W_{t-1}\lambda_{t-1}\right);\ \sum_{j=t-t_1-t_2+1}^{t-1}W_j\lambda_j\geq m\right]+\\
&\qquad\mathbb{E}\left[\pi_t^{LT}\left(W_1\lambda_1,W_2\lambda_2,\ldots,W_{t-1}\lambda_{t-1}\right);\ \sum_{j=t-t_1-t_2+1}^{t-1}W_j\lambda_j<m\right]\\
&\leq\mathbb{E}\left[1;\ \sum_{j=t-t_1-t_2+1}^{t-1}W_j\lambda_j\geq m\right]\quad\text{(using (4))}\\
&=\mathbb{P}\left[\sum_{j=t-t_1-t_2+1}^{t-1}W_j\lambda_j\geq m\right]\\
&\leq\frac{1}{m}\mathbb{E}\left[\sum_{j=t-t_1-t_2+1}^{t-1}W_j\lambda_j\right]\quad\text{(using the Markov inequality)}\\
&\leq\frac{1}{m}\mathbb{E}\left[\sum_{j=t-t_1-t_2+1}^{t-1}\lambda_j\right]\quad\text{(using the independence of } W_t \text{ across } t \text{ and } \mathbb{E}[W_t]=1).
\end{aligned}
$$

Thus, the long-run proportion of periods in which the agent is rewarded satisfies:

$$\begin{aligned}
&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\pi_t^{LT}\left(W_1\lambda_1, W_2\lambda_2,\ldots,W_{t-1}\lambda_{t-1}\right)\right]\\
&\le \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\frac{1}{m}\mathbb{E}\left[\sum_{j=t-t_1-t_2+1}^{t-1}\lambda_j\right]\\
&\le \lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left[\frac{1}{m}\sum_{t=1}^{T}(t_1+t_2-1)\cdot\lambda_t\right]\\
&=\frac{\overline{\lambda}}{m}\cdot(t_1+t_2-1), \quad \text{where} \quad \overline{\lambda}:=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\lambda_t]. \qquad \text{(A.1)}
\end{aligned}$$

Also, note that the long-run proportion of periods in which the agent is rewarded is at most 1. Thus, under the limited-term reward policy, given an effort policy λ, the agent's expected long-run per-period utility satisfies:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[v\cdot\pi_t^{LT}-c\cdot\lambda_t\right]\le v\cdot\min\left[1,\frac{\overline{\lambda}}{m}\cdot(t_1+t_2-1)\right]-c\cdot\overline{\lambda}.$$

Using this, an upper bound on the agent's expected long-run per-period utility under the limited-term reward policy is obtained as:

$$\max_{\overline{\lambda}\ge 0}\quad v\cdot\min\left[1,\left(\frac{\overline{\lambda}}{m}\cdot(t_1+t_2-1)\right)\right]-c\cdot\overline{\lambda}.$$

Note that any choice of $\overline{\lambda}>\frac{m}{t_1+t_2-1}$ is dominated by $\overline{\lambda}=\frac{m}{t_1+t_2-1}$. Therefore, it is sufficient to consider:

$$\max_{0\le\overline{\lambda}\le\frac{m}{t_1+t_2-1}}\quad \overline{\lambda}\cdot\left(\frac{t_1+t_2-1}{m}\cdot v-c\right).$$

The solution to the above optimization problem is as follows: (i) If $r=\frac{v}{c}\ge\frac{m}{t_1+t_2-1}$, then $\overline{\lambda}=\frac{m}{t_1+t_2-1}$. Thus, the desired upper bound in this case is $v-c\cdot\frac{m}{t_1+t_2-1}$. (ii) If $r=\frac{v}{c}<\frac{m}{t_1+t_2-1}$, then $\overline{\lambda}=0$ and the desired upper bound is 0. This completes our proof of Lemma A.1. $\square$

LEMMA A.2. *Assume that the output of the agent is deterministic – that is, given an input effort $\lambda_t\in\mathbb{R}_+$ by the agent in period t, his output is $q_t=\lambda_t$. Under the limited-term reward policy (t_1,t_2,m), the following effort policy achieves the upper bound characterized in Lemma A.1 on the agent's long-run per-period utility: (i) If $r\ge m/(t_1+t_2-1)$, then $\lambda_t=m$ for $t\in\{1,1+(t_1+t_2-1),1+2\cdot(t_1+t_2-1),\ldots\}$, and $\lambda_t=0$ otherwise. (ii) If $r<m/(t_1+t_2-1)$, then $\lambda_t=0$ for all $t\in\mathbb{N}$.*

Proof of Lemma A.2. Consider the case $r\ge m/(t_1+t_2-1)$. Note that, under the limited-term reward policy, if the agent exerts effort $\lambda_t=m$ in any period t and the output is deterministic, then he is rewarded for periods $t+1,\ldots,t+(t_1+t_2-1)$ due to his output in period t. Under the effort policy specified in Lemma A.2, it is easy to see that the agent is rewarded in every period $t\ge 2$, that is, $\pi_t^{LT}=1$ for all $t\ge 2$. Given that $\pi_1^{LT}=0$ (see (4)), the long-run proportion of periods in which the agent is rewarded is

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\pi_t^{LT}=1.$$

Further, the long-run per-period effort exerted by the agent is $m/(t_1+t_2-1)$. Then, the long-run per-period utility of the agent is

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\left[v\cdot\pi_t^{LT}-c\cdot\lambda_t\right]=v-c\cdot m/(t_1+t_2-1),$$

achieving the desired upper bound.

Consider the case $r<m/(t_1+t_2-1)$. Given $m>0$, under the proposed effort policy, the agent is never rewarded and his long-run per-period utility is 0, again achieving the desired upper bound. This completes the proof of Lemma A.2. □

Lemmas A.1 and A.2 complete the proof of Lemma 1. ■

Appendix B: Proofs of Results in Sections 5 and 6

Remark: For a better understanding, the proof of Theorem 2 is provided later, after the proof of Lemma 7.

Proof of Lemma 3. Consider a limited-term reward policy with parameters $t_1\ge t_1'$, $t_2=1$, and $m>0$. Let $U(c,v,\lambda)$ and $U(c,v,\lambda')$ denote the agent's expected long-run per-period utilities if he follows the effort policies $\lambda(r)$ and $\lambda'(r)$, respectively. We wish to show that $U(c,v,\lambda')\ge U(c,v,\lambda)-c\cdot\Delta_2(\epsilon',\delta',r)$.

Consider $r<m/t_1$. Since $\lambda_t'(r)=0$ for all t (from (7)), we have $U(c,v,\lambda')=0$. Also, from Lemma A.1, since $U(c,v,\lambda)\le 0$, it must be the case that $U(c,v,\lambda)=0$. Thus, $U(c,v,\lambda')=U(c,v,\lambda)=0$.

Consider $m/t_1\le r<\frac{m}{(1-\delta')t_1}$. Since $\lambda_t'(r)=0$ for all t (from (7)), we have $U(c,v,\lambda')=0$. Further, from Lemma A.1, $U(c,v,\lambda)\le v-cm/t_1$. Thus, we have:

$$\begin{aligned}
&U(c,v,\lambda)\le v-\frac{cm}{t_1}<\frac{cm}{(1-\delta')t_1}-\frac{cm}{t_1}=U(c,v,\lambda')+c\cdot\left\{\frac{m}{t_1}\left[\frac{1}{(1-\delta')}-1\right]\right\}\\
\implies\quad &U(c,v,\lambda)-c\cdot\left\{\frac{m}{t_1}\left[\frac{1}{(1-\delta')}-1\right]\right\}<U(c,v,\lambda'),
\end{aligned}$$

where the term in the curly bracket is strictly positive and converges to 0 as $\delta'\to 0$.

Consider $r\ge\frac{m}{(1-\delta')t_1}$. From Lemma A.1, $U(c,v,\lambda)\le v-cm/t_1$. We now obtain a lower bound on the long-run proportion of periods in which the agent with value-to-cost ratio r is rewarded if he follows the effort policy $\lambda'(r)$. We have:

$$\begin{aligned}
\mathbb{E}\left[\pi_t^{LT}\right]&=\mathbb{E}\left[\pi_t^{LT};\ \sum_{j=t-t_1}^{t-1}W_j\lambda_j'(r)\ge m\right]+\mathbb{E}\left[\pi_t^{LT};\ \sum_{j=t-t_1}^{t-1}W_j\lambda_j'(r)<m\right]\\
&=\mathbb{P}\left[\sum_{j=t-t_1}^{t-1}W_j\lambda_j'(r)\ge m\right]\quad\text{(using (4) and the fact that } t_2=1)\\
&=\mathbb{P}\left[\sum_{j=t-t_1}^{t-1}\frac{W_j}{t_1}\ge 1-\delta'\right]\quad\left(\because\ \lambda_t'(r)=\tfrac{m}{(1-\delta')t_1}\ \forall t\text{ from (7) under the given case}\right).
\end{aligned}$$

Thus,

$$\overline{\pi}(r;\lambda')=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[\pi_t^{LT}\right]=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{P}\left[\sum_{j=t-t_1}^{t-1}\frac{W_j}{t_1}\ge 1-\delta'\right]$$

$$
\begin{aligned}
&= \lim_{T\to\infty} \frac{1}{T} \left\{ \sum_{t=1}^{t_1} \mathbb{P}\left[\sum_{j=t-t_1}^{t-1} \frac{W_j}{t_1} \geq 1-\delta' \right] + \sum_{t=t_1+1}^{T} \mathbb{P}\left[\sum_{j=t-t_1}^{t-1} \frac{W_j}{t_1} \geq 1-\delta' \right] \right\} \\
&\geq \lim_{T\to\infty} \frac{1}{T} \sum_{t=t_1+1}^{T} \mathbb{P}\left[\sum_{j=t-t_1}^{t-1} \frac{W_j}{t_1} \geq 1-\delta' \right] \\
&= \lim_{T\to\infty} \frac{1}{T} \sum_{t=t_1+1}^{T} \mathbb{P}\left[\sum_{j=1}^{t_1} \frac{W_j}{t_1} \geq 1-\delta' \right] \quad \text{(since } W_t \text{ is i.i.d. across } t\text{)} \\
&= \mathbb{P}\left[\sum_{j=1}^{t_1} \frac{W_j}{t_1} \geq 1-\delta' \right] \\
&\geq \mathbb{P}\left[\left| \sum_{j=1}^{t_1} \frac{W_j}{t_1} - 1 \right| \leq \delta' \right] \\
&> 1-\epsilon' \quad \text{(using (6)).}
\end{aligned} \tag{B.1}
$$

Thus, the expected long-run per-period utility of the agent whose value-to-cost ratio is r and follows the effort policy $\lambda'(r)$ is:

$$
\begin{aligned}
U(c,v,\lambda') &= v \cdot \overline{\pi}(r;\lambda') - c \cdot \overline{\lambda}'(r) \\
&> v(1-\epsilon') - \frac{cm}{(1-\delta')t_1} \\
&= v - \frac{cm}{t_1} - v\epsilon' - \frac{cm\delta'}{(1-\delta')t_1} \\
&\geq U(c,v,\lambda) - c \cdot \left[r\epsilon' + \frac{m\delta'}{(1-\delta')t_1} \right],
\end{aligned} \tag{B.2}
$$

where the term in the square bracket is strictly postive and converges to 0 as $\epsilon' \to 0$ and $\delta' \to 0$.

Together, the above three cases complete the proof of Lemma 3. ■

Proof of Lemma 4. Note that $r = \frac{m}{(1-\delta')t_1}$. Using (B.2), (9) can be written as:

$$
0 \;\leq\; r \cdot [\overline{\pi}(r;\lambda) - \overline{\pi}(r;\lambda')] - \left[\overline{\lambda}(r) - \overline{\lambda}'(r) \right] \;\leq\; \left[r\epsilon' + \frac{m\delta'}{(1-\delta')t_1} \right]. \tag{B.3}
$$

Further, we have $\overline{\lambda}'(r) = \frac{m}{(1-\delta')t_1}$ (see (7)). Using this, along with $\overline{\pi}(r;\lambda') > 1-\epsilon'$ from (B.1), (10) can be simplified as:

$$
\begin{aligned}
\overline{\lambda}(r) &\;\geq\; \frac{m}{t_1} \cdot \overline{\pi}(r;\lambda) \\
\Leftrightarrow \quad \left[\overline{\lambda}(r) - \overline{\lambda}'(r) \right] &\;\geq\; \frac{m}{t_1} \cdot [\overline{\pi}(r;\lambda) - \overline{\pi}(r;\lambda')] + \frac{m}{t_1} \cdot \overline{\pi}(r;\lambda') - \overline{\lambda}'(r) \\
\implies \quad \left[\overline{\lambda}(r) - \overline{\lambda}'(r) \right] &\;\geq\; \frac{m}{t_1} \cdot [\overline{\pi}(r;\lambda) - \overline{\pi}(r;\lambda')] + \frac{m(1-\epsilon')}{t_1} - \frac{m}{(1-\delta')t_1}.
\end{aligned} \tag{B.4}
$$

Also, since $\delta' \in (0,1)$, we have $r > m/t_1$ and $r \geq c_0/v_0$ (by assumption; see Section 3).

To obtain a lower bound on Q_1, we solve the following linear program:

$$
\min \quad \left[v_0 \cdot \overline{\lambda}(r) - c_0 \cdot \overline{\pi}(r;\lambda) \right] - \left[v_0 \cdot \overline{\lambda}'(r) - c_0 \cdot \overline{\pi}(r;\lambda') \right] \quad \text{s.t.} \quad \text{(B.3)–(B.4).}
$$

The solution of the above linear program is summarized by the two cases below.

Case 1: $m/t_1 < c_0/v_0$. Under the optimal solution of the linear program, the following two constraints are binding:

$$r \cdot [\overline{\pi}(r;\lambda) - \overline{\pi}(r;\lambda')] - \left[\overline{\lambda}(r) - \overline{\lambda}'(r)\right] \;\leq\; \left[r\epsilon' + \frac{m\delta'}{(1-\delta')t_1}\right], \quad \text{and,}$$

$$\left[\overline{\lambda}(r) - \overline{\lambda}'(r)\right] \;\geq\; \frac{m}{t_1} \cdot [\overline{\pi}(r;\lambda) - \overline{\pi}(r;\lambda')] + \frac{m(1-\epsilon')}{t_1} - \frac{m}{(1-\delta')t_1}.$$

Solving the binding constraints at equality and performing algebraic simplifications, we have the following lower bound on Q_1:

$$Q_1 = \left[v_0 \cdot \overline{\lambda}(r) - c_0 \cdot \overline{\pi}(r;\lambda)\right] - \left[v_0 \cdot \overline{\lambda}'(r) - c_0 \cdot \overline{\pi}(r;\lambda')\right] \geq -v_0 \cdot \left[\frac{m\delta'}{(1-\delta')t_1} + \frac{c_0\epsilon'}{v_0}\right].$$

Thus, using $m = (1-\delta')t_1 r$, we have

$$Q_1 \geq -v_0 \cdot \left[\delta' r + \frac{c_0\epsilon'}{v_0}\right], \tag{B.5}$$

where the term in the right-hand side is strictly negative and converges to 0 as $\epsilon' \to 0$ and $\delta' \to 0$.

Case 2: $m/t_1 \geq c_0/v_0$. Under the optimal solution of the linear program, the following two constraints are binding:

$$r \cdot [\overline{\pi}(r;\lambda) - \overline{\pi}(r;\lambda')] - \left[\overline{\lambda}(r) - \overline{\lambda}'(r)\right] \geq 0, \quad \text{and,}$$

$$\left[\overline{\lambda}(r) - \overline{\lambda}'(r)\right] \;\geq\; \frac{m}{t_1} \cdot [\overline{\pi}(r;\lambda) - \overline{\pi}(r;\lambda')] + \frac{m(1-\epsilon')}{t_1} - \frac{m}{(1-\delta')t_1}.$$

Solving the binding constraints at equality and performing algebraic simplifications, we obtain the following lower bound on Q_1:

$$\begin{aligned} Q_1 &= \left[v_0 \cdot \overline{\lambda}(r) - c_0 \cdot \overline{\pi}(r;\lambda)\right] - \left[v_0 \cdot \overline{\lambda}'(r) - c_0 \cdot \overline{\pi}(r;\lambda')\right] \\ &\geq -v_0 \cdot \frac{m}{t_1} \cdot \left[1 + \frac{m/t_1 - c_0/v_0}{r - m/t_1}\right] \cdot \left[\frac{1}{1-\delta'} - (1-\epsilon')\right]. \end{aligned}$$

Thus, using $m = (1-\delta')t_1 r$, we have

$$Q_1 \geq -v_0 \cdot (1-\delta')r \cdot \left[\frac{r - c_0/v_0}{\delta' r}\right] \cdot \left[\frac{1}{1-\delta'} - (1-\epsilon')\right], \tag{B.6}$$

where the term in the right-hand side is strictly negative and converges to 0 as $\epsilon' \to 0$ and $\delta' \to 0$. The above two cases complete the proof of Lemma 4. ■

Proof of Lemma 5. Since $r = \frac{m}{(1-\delta')t_1}$, using (8), we have:

$$\begin{aligned} Q_2 &= \left[v_0 \cdot \overline{\lambda}'(r) - c_0 \cdot \overline{\pi}(r;\lambda')\right] \\ &\geq v_0 \cdot \frac{m}{(1-\delta')t_1} - c_0 \quad (\text{using } \overline{\pi}(r;\lambda') \leq 1 \text{ and } (7)) \\ &= v_0 \cdot r - c_0 \\ &\geq OPT \quad (\text{using } (3)), \end{aligned}$$

completing the proof of Lemma 5. ■

Proof of Lemma 6. Using (A.1) and substituting $t_2 = 1$, we obtain the desired result. ■

Proof of Theorem 4. Fix an arbitrary $\delta' \in (0,1)$. Consider the limited-term reward policy with parameters $t_1 = \bar{t}_1$, $t_2 = 1$, and $m = (1-\delta') \cdot t_1 \cdot r$. By leveraging the proof of Theorem 3, we will first obtain a lower bound on the principal's expected long-run per-period utility under this policy. This, in turn, will also establish the same lower bound on the principal's objective value under an *optimal* policy within the class $LT(\bar{t}_1, 1)$. Then, we will establish the desired monotonic and asymptotic results.

Under the given limited-term reward policy, let $\lambda(r)$ be an optimal effort policy for the agent with value-to-cost ratio r, that is, one that solves (IC). Let $\lambda'(r)$ be the effort policy defined below. For any $t \in \mathbb{N}$,

$$\lambda'_t(r) = \begin{cases} \frac{m}{(1-\delta')t_1} & r \geq \frac{m}{(1-\delta')t_1} \\ 0 & \text{otherwise.} \end{cases}$$

Let $\overline{\pi}(r;\lambda)$ (resp., $\overline{\pi}(r;\lambda')$) be the expected long-run proportion of periods in which the agent is rewarded, and $\overline{\lambda}(r)$ (resp., $\overline{\lambda}'(r)$) be the expected long-run per-period effort exerted by the agent, given that he follows the effort policy $\lambda(r)$ (resp., $\lambda'(r)$). Then, the principal's expected long-run per-period utility under the limited-term reward policy is given as:

$$\begin{aligned} &\left[v_0 \cdot \overline{\lambda}(r) - c_0 \cdot \overline{\pi}(r;\lambda)\right] \\ &= \underbrace{\left[v_0 \cdot \overline{\lambda}(r) - c_0 \cdot \overline{\pi}(r;\lambda)\right] - \left[v_0 \cdot \overline{\lambda}'(r) - c_0 \cdot \overline{\pi}(r;\lambda')\right]}_{Q_1} \;+\; \underbrace{\left[v_0 \cdot \overline{\lambda}'(r) - c_0 \cdot \overline{\pi}(r;\lambda')\right]}_{Q_2}. \end{aligned} \tag{B.7}$$

We consider the following cases:

<u>Case 1: $m/t_1 < c_0/v_0$.</u> Define $\epsilon' := \min\{1, \sigma^2/(\bar{t}_1\delta'^2)\}$. Using Chebyshev's inequality, we have:

$$\mathbb{P}\left[\left|\sum_{j=1}^{\bar{t}_1} \frac{W_j}{\bar{t}_1} - 1\right| \leq \delta'\right] \geq \max\left\{1 - \frac{\sigma^2}{\bar{t}_1\delta'^2},\ 0\right\} = 1 - \frac{\sigma^2}{\bar{t}_1\delta'^2} - \min\left\{1 - \frac{\sigma^2}{\bar{t}_1\delta'^2},\ 0\right\} = 1 - \epsilon'. \tag{B.8}$$

We note that (B.8) is analogous to (6). Following arguments similar to those in the proof of Theorem 3, and using (B.5), we have:

$$Q_1 \geq -v_0 \cdot \left[\delta' r + \frac{c_0\epsilon'}{v_0}\right] = -Q_{1,a}(\epsilon', \delta'),$$

where we note that $Q_{1,a}(\epsilon', \delta')$ is strictly positive and increasing in ϵ' for any fixed $\delta' \in (0,1)$.

<u>Case 2: $m/t_1 \geq c_0/v_0$.</u> Following arguments similar to those in the proof of Theorem 3, and using (B.6), we have:

$$Q_1 \geq -v_0 \cdot (1-\delta') r \cdot \left[\frac{r - c_0/v_0}{\delta' r}\right] \cdot \left[\frac{1}{1-\delta'} - (1-\epsilon')\right] = -Q_{1,b}(\epsilon', \delta').$$

Clearly, $Q_{1,b}(\epsilon', \delta')$ is strictly positive and increasing in ϵ' for any fixed $\delta' \in (0,1)$.

Using the above cases, we obtain a lower bound on Q_1 as follows:

$$Q_1 \geq -\max\left\{Q_{1,a}(\epsilon', \delta'), Q_{1,b}(\epsilon', \delta')\right\}.$$

We define

$$\epsilon(\bar{t}_1, \sigma; \delta') := \max\left\{Q_{1,a}(\epsilon', \delta'), Q_{1,b}(\epsilon', \delta')\right\}, \quad \text{where we recall } \; \epsilon' = \min\{1, \sigma^2/(\bar{t}_1\delta'^2)\}. \tag{B.9}$$

Since the above expression is increasing in ϵ', and ϵ' is decreasing in $\bar{t}_1$ and increasing in $\sigma > 0$, we have that $\epsilon(\bar{t}_1, \sigma; \delta')$ is decreasing in $\bar{t}_1$ and increasing in $\sigma > 0$.

Further, by following similar steps as in the proof of Lemma 5, we also have $Q_2 \geq OPT$. Thus, from (B.7), we have:

$$\left[v_0 \cdot \overline{\lambda}(R) - c_0 \cdot \overline{\pi}(R; \lambda)\right] \geq OPT - \epsilon(\bar{t}_1, \sigma; \delta'),$$

establishing the first part of Theorem 4. The second part of Theorem 4, that is, the monotonicity of $\epsilon(\bar{t}_1, \sigma; \delta')$ with respect to $\bar{t}_1$ and σ, has already been established in (B.9) above.

Since $\delta' \in (0,1)$ is arbitrary, to establish the stated asymptotic property of $\epsilon(\bar{t}_1, \sigma; \delta')$ with respect to $\bar{t}_1$, we set $\delta' = 1/(1+\bar{t}_1)^{1/4}$. Since $\bar{t}_1 \geq 1$, we have $\delta' \in (0,1)$. Note that, as $\bar{t}_1 \to \infty$, we have $\delta' \to 0$ and $\epsilon' \to 0$. Consequently, $Q_{1,a}(\epsilon', \delta') \to 0$, and $Q_{1,b}(\epsilon', \delta') \to 0$, and therefore, from (B.9), $\epsilon(\bar{t}_1, \sigma; \delta') \to 0$. This establishes the third part of Theorem 4, completing our proof. ∎

Remark: The proof of Theorem 5 is provided at the end of this appendix.

Proof of Lemma 7. Consider the score-based reward policy with parameter θ. We first obtain an upper bound on the expected long-run per-period utility of the agent for any arbitrary effort policy $\lambda \in \Lambda$ (Lemma B.1). Then, we identify an effort policy which achieves this bound, thus establishing its optimality (Lemma B.3). To derive the latter result, we also need an intermediate result that we establish in Lemma B.2.

LEMMA B.1. *Under the score-based reward policy with parameter θ, the expected long-run per-period utility of the agent is bounded from above as follows: (i) If $r \geq 1/\theta$, then the upper bound is $v - c/\theta$. (ii) If $r < 1/\theta$, then the upper bound is 0.*

Proof of Lemma B.1. Using (1), under the score-based reward policy with parameter θ, the agent's expected long-run per-period utility is:

$$\begin{aligned}
&\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[v \cdot \pi_t^{SB} - c \cdot \lambda_t\right] = v \cdot \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[\pi_t^{SB}\right] - c\overline{\lambda}, \quad \text{where} \quad \overline{\lambda} := \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\lambda_t] \\
&= v \cdot \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[S_t - S_{t+1} + \theta W_t \lambda_t\right] - c\overline{\lambda} \quad \text{(using (12))} \\
&= v \cdot \left(-\lim_{T\to\infty} \frac{\mathbb{E}[S_{T+1}]}{T} + \theta\overline{\lambda}\right) - c\overline{\lambda} \quad \text{(using } S_1 = 0\text{, independence of } W_t \text{ across } t \text{ and } \mathbb{E}[W_t] = 1\ \forall t) \\
&= -v \cdot \lim_{T\to\infty} \frac{\mathbb{E}[S_{T+1}]}{T} + \overline{\lambda} \cdot (v\theta - c) \qquad \text{(B.10)} \\
&\leq \overline{\lambda} \cdot (v\theta - c),
\end{aligned}$$

where the inequality follows from the fact that $S_t \geq 0$ for all t (see (12)). Further, since $\pi_t^{SB} \leq 1$ for all t, the agent's expected long-run per-period utility is also upper bounded by $v - c\overline{\lambda}$. Combining these two observations, we have

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[v \cdot \pi_t^{SB} - c \cdot \lambda_t\right] \leq v \cdot \min(\theta\overline{\lambda}, 1) - c\overline{\lambda}.$$

Thus, an upper bound on the agent's expected long-run per-period utility can be obtained by maximizing $v \cdot \min(\theta\overline{\lambda}, 1) - c\overline{\lambda}$ with respect to $\overline{\lambda}$. Note that any choice of $\overline{\lambda} > 1/\theta$ is dominated by $\overline{\lambda} = 1/\theta$. Therefore, it is sufficient to consider:

$$\max_{0 \le \overline{\lambda} \le 1/\theta} \quad \overline{\lambda} \cdot (v\theta - c).$$

The solution to the above optimization problem is as follows: (i) If $r = v/c \ge 1/\theta$, then $\overline{\lambda} = 1/\theta$. Thus, the desired upper bound in this case is $v - c/\theta$. (ii) If $r = v/c < 1/\theta$, then $\overline{\lambda} = 0$ and the desired upper bound is 0. This completes our proof of Lemma B.1. □

LEMMA B.2. *Under the score-based reward policy with parameter θ, consider the effort policy $\lambda_t = 1/\theta$ for all $t \in \mathbb{N}$. Then, we have:*

$$\lim_{T\to\infty} \frac{\mathbb{E}[S_{T+1}]}{T} = 0.$$

Proof of Lemma B.2. Define a sequence $\{X_t, t \in \mathbb{N}\}$ as follows:

$$X_{t+1} = X_t - \min(X_t, 1) + \theta W_t \lambda_t, \tag{B.11}$$

where $X_1 = 0$. We use the notation $x^+ := \max(x, 0)$ for any $x \in \mathbb{R}$. Using the identity $a + b = \min(a, b) + \max(a, b)$ in (B.11), we have $X_{t+1} = (X_t - 1)^+ + \theta W_t \lambda_t$. Since $W_t \in \mathbb{R}_+$ and $\lambda_t \in \mathbb{R}_+$, we have $X_t \ge 0$ for all $t \in \mathbb{N}$. Define another sequence $\{Y_t, t \in \mathbb{N}\}$ where $Y_t = (X_t - 1)^+$. Then, we have:

$$Y_{t+1} = (X_{t+1} - 1)^+ = (Y_t + \theta W_t \lambda_t - 1)^+.$$

The sequence $\{Y_t, t \in \mathbb{N}\}$ can be viewed as a sequence of the waiting times in a single-server first-come-first-serve queue. Note that, using the assumption that $\lambda_t = 1/\theta$ for all t and $\mathbb{E}[W_t] = 1$ for all t, we have $\mathbb{E}[\theta W_t \lambda_t] = \theta \cdot 1/\theta = 1$. Then, from Theorem 4 of Loynes (1962), we have:

$$\lim_{T\to\infty} \frac{\mathbb{E}[Y_T]}{T} = 0. \tag{B.12}$$

Thus, we have:

$$\begin{aligned}
\lim_{T\to\infty} \frac{\mathbb{E}[X_{T+1}]}{T} &= \lim_{T\to\infty} \frac{\mathbb{E}[(X_T - 1)^+ + \theta W_T \lambda_T]}{T} \\
&= \lim_{T\to\infty} \frac{\mathbb{E}[Y_T + \theta W_T \lambda_T]}{T} \\
&= \underbrace{\lim_{T\to\infty} \frac{\mathbb{E}[Y_T]}{T}}_{=0 \quad ((\text{B.12}))} + \underbrace{\lim_{T\to\infty} \frac{\mathbb{E}[\theta W_T \lambda_T]}{T}}_{=0 \quad (\because \mathbb{E}[W_T] = 1;\, \lambda_T = 1/\theta)} \\
&= 0.
\end{aligned} \tag{B.13}$$

We now exploit the connection between the sequences $\{X_t, t \in \mathbb{N}\}$ and $\{S_t, t \in \mathbb{N}\}$ to prove Lemma B.2. First, we show

$$X_t \le S_t \le X_t + 1 \quad \forall t \in \mathbb{N}. \tag{B.14}$$

We use induction to prove (B.14). Since $S_1 = X_1 = 0$, (B.14) holds for $t = 1$. Suppose, for the induction hypothesis, that (B.14) holds for some $t \ge 1$. We consider the following three (exhaustive) cases:

(i) $\underline{X_t \geq 1}$: Then, $\min(X_t, 1) = 1$. Further, by the induction hypothesis, $1 \leq X_t \leq S_t$. Thus, using (12), $\pi_t^{SB} = 1$. Thus, we have:

$$\begin{aligned} X_{t+1} &= X_t - \min(X_t, 1) + \theta W_t \lambda_t \\ &\leq S_t - 1 + \theta W_t \lambda_t \\ &= S_t - \pi_t^{SB} + \theta W_t \lambda_t = S_{t+1}. \end{aligned}$$

Also, since $S_t \leq X_t + 1$ by the induction hypothesis, we have

$$\begin{aligned} S_{t+1} &= S_t - \pi_t^{SB} + \theta W_t \lambda_t \\ &= S_t - 1 + \theta W_t \lambda_t \\ &\leq (X_t + 1) - \min(X_t, 1) + \theta W_t \lambda_t = X_{t+1} + 1. \end{aligned}$$

Thus, we have established $X_{t+1} \leq S_{t+1} \leq X_{t+1} + 1$ for the case $X_t \geq 1$.

(ii) $\underline{S_t < 1}$. Then, using (12), $\pi_t^{SB} = 0$. Further, by the induction hypothesis, $X_t \leq S_t < 1$; thus, $\min(X_t, 1) = X_t \geq 0 = \pi_t^{SB}$. As a result, we have

$$\begin{aligned} X_{t+1} &= X_t - \min(X_t, 1) + \theta W_t \lambda_t \\ &\leq S_t - \pi_t^{SB} + \theta W_t \lambda_t = S_{t+1}. \end{aligned}$$

Also, we have

$$\begin{aligned} S_{t+1} &= S_t - \pi_t^{SB} + \theta W_t \lambda_t \\ &< 1 + \theta W_t \lambda_t = 1 + X_t - \min(X_t, 1) + \theta W_t \lambda_t \\ &= 1 + X_{t+1}. \end{aligned}$$

Thus, we have established $X_{t+1} \leq S_{t+1} \leq X_{t+1} + 1$ for the case $S_t < 1$.

(iii) $\underline{X_t < 1 \leq S_t}$. Then, $\pi_t^{SB} = 1$ from (12) and $\min(X_t, 1) = X_t$. Since $S_t \geq 1 = \pi_t^{SB}$, we have

$$\begin{aligned} X_{t+1} &= X_t - \min(X_t, 1) + \theta W_t \lambda_t \\ &= \theta W_t \lambda_t \leq S_t - \pi_t^{SB} + \theta W_t \lambda_t = S_{t+1}. \end{aligned}$$

Also, since $S_t \leq X_t + 1$ (induction hypothesis) and $X_t < 1$, we have $S_t < 2$. Thus,

$$\begin{aligned} S_{t+1} &= S_t - \pi_t^{SB} + \theta W_t \lambda_t \\ &= S_t - 1 + \theta W_t \lambda_t \\ &< 1 + \theta W_t \lambda_t = 1 + X_t - \min(X_t, 1) + \theta W_t \lambda_t \\ &= 1 + X_{t+1}. \end{aligned}$$

Thus, from the above three cases, we conclude the induction argument and hence the proof of (B.14).

Using (B.14), we have:

$$\begin{aligned} & \lim_{T\to\infty} \frac{\mathbb{E}[X_{T+1}]}{T} \;\le\; \lim_{T\to\infty} \frac{\mathbb{E}[S_{T+1}]}{T} \;\le\; \lim_{T\to\infty} \frac{1+\mathbb{E}[X_{T+1}]}{T} \\ \Longrightarrow \quad & 0 \;\le\; \lim_{T\to\infty} \frac{\mathbb{E}[S_{T+1}]}{T} \;\le\; 0 \qquad \text{(using (B.13))} \\ \Longrightarrow \quad & \lim_{T\to\infty} \frac{\mathbb{E}[S_{T+1}]}{T} = 0, \end{aligned}$$

thus achieving the desired result and completing the proof of Lemma B.2. □

LEMMA B.3. *Under the score-based reward policy with parameter θ, the following effort policy achieves the upper bound characterized in Lemma B.1 on the agent's expected long-run per-period utility: (i) If $r \ge 1/\theta$, then $\lambda_t = 1/\theta$ for all t. (ii) If $r < 1/\theta$, then $\lambda_t = 0$ for all t.*

Proof of Lemma B.3. Consider the case $r \ge 1/\theta$. Using (B.10), under the score-based reward policy with parameter θ, the expected long-run per-period utility of the agent when he exerts the effort $\lambda_t = 1/\theta$ in every period t is:

$$\begin{aligned} & -v \cdot \lim_{T\to\infty} \frac{\mathbb{E}[S_{T+1}]}{T} + \overline{\lambda} \cdot (v\theta - c) \\ & = \overline{\lambda} \cdot (v\theta - c) \qquad \text{(using Lemma B.2)} \\ & = v - c/\theta, \end{aligned}$$

achieving the desired upper bound.

Consider the case $r < 1/\theta$. Under the score-based reward policy with parameter θ, the expected long-run per-period utility of the agent when he exerts the effort $\lambda_t = 0$ in every period t is:

$$\begin{aligned} & \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[v \cdot \pi_t^{SB} - c \cdot \lambda_t\right] \\ & = 0. \qquad (\because\ \lambda_t = 0\ \forall t \text{ implies } q_t = w_t \lambda_t = 0,\ S_t = 0 \text{ and } \pi_t^{SB} = 0\ \forall t) \end{aligned}$$

The above two cases complete the proof of Lemma B.3. □

Lemmas B.1 and B.3 complete the proof of Lemma 7. ■

Proof of Theorem 2. Consider the following instance of problem $\mathcal{P}$: Suppose that, for every $t \in \mathbb{N}$, we have $W_t = 2$ with probability $1/2$, and $W_t = 0$ otherwise. Further, assume that $OPT_B = v_0 r - c_0 > 0$, where OPT_B is obtained in (3).

From the proof of Theorem 6, we know that, under an optimal score-based reward policy, the principal's expected long-run per-period utility equals OPT_B. This, together with $OPT \le OPT_B$ (from (3)), implies that $OPT = OPT_B$. Thus, in problem $\mathcal{P}$, the principal's expected long-run per-period utility under any limited-term reward policy can be equal to OPT (or equivalently, OPT_B) *only if* that reward policy also solves the relaxation $\mathcal{P}_B$ (defined in Section 3) of $\mathcal{P}$. Therefore, to prove Theorem 2, under the given instance of problem $\mathcal{P}$, we show that *every* limited-term reward policy is *sub-optimal* for problem $\mathcal{P}_B$.

Consider an arbitrary limited-term reward policy with parameters (t_1, t_2, m), where, from Section 4, we know that $t_1 \geq 1$, $t_2 \geq 1$ and $m > 0$. Under this reward policy, let $\lambda^{LT}(r)$ be an optimal effort policy for the agent whose value-to-cost ratio is r. Correspondingly, let $\overline{\pi}^{LT}(r)$ and $\overline{\lambda}^{LT}(r)$ be the expected long-run proportion of periods in which the agent is rewarded and the expected long-run per-period effort exerted by the agent, respectively. Define $y^{LT}(r) = r\overline{\pi}^{LT}(r) - \overline{\lambda}^{LT}(r)$ for all $r \in \mathbb{R}_+$.

We have the following result.

Lemma B.4. *Under the limited-term reward policy with parameters (t_1, t_2, m), for any $r \in \mathbb{R}_+$, we have $\overline{\pi}^{LT}(r) \leq 1 - (1/2)^{t_1+t_2-1}$.*

Proof of Lemma B.4. Fix an arbitrary $r \in \mathbb{R}_+$. Under the given limited-term reward policy, the agent with value-to-cost ratio r exerts effort according to an optimal effort policy $\lambda^{LT}(r)$. Recall from (4) that, if the agent is rewarded in period t, $\exists k \in \{t - t_2, \ldots, t-1\}$ such that $\sum_{j=k-t_1+1}^{k} q_j \geq m$ where $q_j = w_j \lambda_j^{LT}(r)$ for all $j \in \mathbb{N}$. Thus, if the agent receives the reward in period t (that is, $\pi_t^{LT}(r) = 1$), then his cumulative output over periods $t - (t_1 + t_2 - 1)$ to $t - 1$ is at least m. That is,

$$\pi_t^{LT}(r) = 1 \implies \sum_{j=t-(t_1+t_2-1)}^{t-1} q_j \geq m. \tag{B.15}$$

Then,

$$\begin{aligned}
&\mathbb{E}\Big[\pi_t^{LT}(r)\Big] \\
&= \mathbb{E}\Big[\pi_t^{LT}(r);\ \exists j \in \{t-(t_1+t_2-1), \ldots, t-1\}:\ W_j \neq 0\Big] \ + \ \underbrace{\mathbb{E}\Big[\pi_t^{LT}(r);\ W_j = 0\ \forall j \in \{t-(t_1+t_2-1), \ldots, t-1\}\Big]}_{=0 \ \ (\text{since } \pi_t^{LT}(r)=0 \text{ by (B.15)})} \\
&\leq \mathbb{E}\Big[1;\ \exists j \in \{t-(t_1+t_2-1), \ldots, t-1\}:\ W_j \neq 0\Big] \qquad (\because \pi_t^{LT}(r) \leq 1\ \forall t) \\
&= \mathbb{P}\Big[\exists j \in \{t-(t_1+t_2-1), \ldots, t-1\}:\ W_j \neq 0\Big] \\
&= 1 - \mathbb{P}\Big[W_j = 0\ \forall j \in \{t-(t_1+t_2-1), \ldots, t-1\}\Big] \\
&= 1 - \left(\frac{1}{2}\right)^{\min[t-1, t_1+t_2-1]},
\end{aligned} \tag{B.16}$$

where the last equality follows by considering both cases $t < 1 + (t_1 + t_2 - 1)$ and $t \geq 1 + (t_1 + t_2 - 1)$. Thus, we have:

$$\begin{aligned}
\overline{\pi}^{LT}(r) &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\Big[\pi_t^{LT}(r)\Big] \\
&= \lim_{T\to\infty} \frac{1}{T} \left\{ \sum_{t=1}^{t_1+t_2-1} \mathbb{E}\Big[\pi_t^{LT}(r)\Big] + \sum_{t=t_1+t_2}^{T} \mathbb{E}\Big[\pi_t^{LT}(r)\Big] \right\} \\
&= \lim_{T\to\infty} \frac{1}{T} \left\{ \sum_{t=2}^{t_1+t_2-1} \mathbb{E}\Big[\pi_t^{LT}(r)\Big] + \sum_{t=t_1+t_2}^{T} \mathbb{E}\Big[\pi_t^{LT}(r)\Big] \right\} \quad (\because \pi_1^{LT}(r) = 0;\ \text{see (4)}) \\
&\leq \lim_{T\to\infty} \frac{1}{T} \left\{ \sum_{t=2}^{t_1+t_2-1} \left[1 - \left(\frac{1}{2}\right)^{t-1}\right] \ + \ \sum_{t=t_1+t_2}^{T} \left[1 - \left(\frac{1}{2}\right)^{t_1+t_2-1}\right] \right\} \quad (\text{using (B.16)})
\end{aligned}$$

$$
\begin{aligned}
&= \underbrace{\lim_{T\to\infty}\frac{1}{T}\sum_{t=2}^{t_1+t_2-1}\left[1-\left(\frac{1}{2}\right)^{t-1}\right]}_{=0\ \text{(since } t_1\in\mathbb{N} \text{ and } t_2\in\mathbb{N} \text{ are fixed)}} + \lim_{T\to\infty}\frac{1}{T}\sum_{t=t_1+t_2}^{T}\left[1-\left(\frac{1}{2}\right)^{t_1+t_2-1}\right] \\
&= \lim_{T\to\infty}\frac{1}{T}\sum_{t=t_1+t_2}^{T}\left[1-\left(\frac{1}{2}\right)^{t_1+t_2-1}\right] \\
&= 1-\left(\frac{1}{2}\right)^{t_1+t_2-1}.
\end{aligned}
$$

This completes the proof of Lemma B.4. □

Recall from Section 3 that problem $\mathcal{P}_B$ can be expressed as

$$
\begin{aligned}
OPT_B \ := \ \max_{\substack{\pi\in\Pi\\ \lambda\in\Lambda}} \quad & \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\Big[v_0\cdot W_t\cdot\lambda_t \ - \ c_0\cdot\pi_t\Big] \\
\text{s.t.} \quad & U(\lambda \mid \pi) \ \geq \ 0. \qquad\qquad (IR)
\end{aligned}
$$

Since the limited-term reward policy π^{LT} with parameters (t_1, t_2, m) is feasible to $\mathcal{P}$, it must also be feasible to the relaxation $\mathcal{P}_B$ of $\mathcal{P}$. Consequently, the principal's objective value in problem $\mathcal{P}_B$ under the given limited-term reward policy is:

$$
\begin{aligned}
& v_0\overline{\lambda}^{LT}(r) - c_0\cdot\overline{\pi}^{LT}(r) \\
& \le (v_0 r - c_0)\cdot\overline{\pi}^{LT}(r) \quad (\text{using } (IR)) \\
& \le \left[1-\left(\frac{1}{2}\right)^{t_1+t_2-1}\right]\cdot(v_0 r - c_0) \quad (\text{using Lemma B.4}) \\
& < OPT_B \quad (\because\ t_1\ge 1, t_2\ge 1, OPT_B > 0).
\end{aligned}
$$

Together, the above set of inequalities imply that $\{\overline{\pi}^{LT}(\cdot), \overline{\lambda}^{LT}(\cdot)\}$ cannot be optimal for $\mathcal{P}_B$. Thus, for the given instance of $\mathcal{P}$, the principal's objective cannot equal OPT_B (and hence OPT) for *any* limited-term reward policy. This completes the proof of Theorem 2. ■

Proof of Theorem 5. Consider an instance of problem $\mathcal{P}$ in which W_t takes the value $(1+\sigma^2)$ with probability $1/(1+\sigma^2)$ and 0 otherwise, where $\sigma > 0$ is the standard deviation of W_t. Further, assume that $OPT_B = v_0 r - c_0 > 0$, where OPT_B is obtained in (3). Fix $\bar{t}_1 = 1$. Since, by definition, $t_1 \ge 1$ (see Section 4), the requirements $t_1 \le \bar{t}_1 = 1$ and $t_1 \ge 1$, together imply that $t_1 = 1$. For the given instance, we show that, under any reward policy in the class $LT(1,1)$, the principal's expected long-run per-period utility is at most $OPT - \eta$, where $\eta > 0$. This will establish Theorem 5.

First, given $t_1 = t_2 = 1$, the agent's optimal effort policy under any reward policy from the class $LT(1,1)$ can be obtained by solving a single-period problem. Subsequently, we characterize the optimal reward policy within the class $LT(1,1)$ for the principal by optimizing the value of m.

Fix an arbitrary $m > 0$. Suppose that the agent exerts an effort λ_t in any period t. Then, his expected utility from this effort is:

$$
v\mathbb{P}[W_t\lambda_t \ge m] - c\lambda_t
$$

$$= v\,\underbrace{\mathbb{P}[W_t\lambda_t \geq m \mid W_t = 0]}_{=0 \quad (\because m>0)}\mathbb{P}[W_t = 0] + v\mathbb{P}[W_t\lambda_t \geq m \mid W_t = 1+\sigma^2]\mathbb{P}[W_t = 1+\sigma^2] - c\lambda_t$$

$$= \begin{cases} -c\lambda_t & \lambda_t \in [0, m/(1+\sigma^2)) \\ v/(1+\sigma^2) - c\lambda_t & \lambda_t \geq m/(1+\sigma^2). \end{cases}$$

Thus, the agent's expected utility is piecewise decreasing in λ_t and optimized at either $\lambda_t = 0$ or $\lambda_t = m/(1+\sigma^2)$, depending on the value of r. Note that $\lambda_t = m/(1+\sigma^2)$ is optimal if $v/(1+\sigma^2) - cm/(1+\sigma^2) \geq 0$, that is, $r \geq m$. Otherwise, $\lambda_t = 0$ is optimal.

Thus, the agent's optimal effort policy is given as follows: For any $t \geq 1$, we have $\lambda_t(r) = m/(1+\sigma^2)$ if $r \geq m$, and $\lambda_t(r) = 0$ otherwise. If $r \geq m$, then the agent receives the reward in any given period with probability $\mathbb{P}[W_t\lambda_t \geq m] = \mathbb{P}[W_t \geq 1+\sigma^2] = 1/(1+\sigma^2)$. If $r < m$, then the agent is never rewarded. Thus, the expected long-run proportion of periods in which the agent is rewarded is:

$$\overline{\pi}(r) = \begin{cases} 1/(1+\sigma^2) & r \geq m \\ 0 & r < m. \end{cases}$$

Thus, the principal's expected long-run per-period utility is given as:

$$v_0\overline{\lambda}(r) - c_0\overline{\pi}(r) = \begin{cases} v_0\frac{m}{1+\sigma^2} - c_0\frac{1}{1+\sigma^2} & r \geq m \\ 0 & r < m. \end{cases} \tag{B.17}$$

Clearly, the principal's expected utility is maximized at $m = r$. Substituting this in (B.17), the principal's expected long-run per-period utility under the optimal policy from the class of policies $LT(1,1)$ is given as:

$$\begin{aligned} v_0\overline{\lambda}(r) - c_0\overline{\pi}(r) &= \frac{1}{1+\sigma^2}\,(v_0 r - c_0) \\ &= \frac{OPT_B}{1+\sigma^2} \quad \text{(using (3))} \\ &= \frac{OPT}{1+\sigma^2} \quad \text{(since } OPT = OPT_B \text{ from the proof of Theorem 6)} \\ &= OPT - \eta, \end{aligned}$$

where $\eta := OPT \cdot \sigma^2/(1+\sigma^2)$ is strictly positive because $OPT = OPT_B > 0$ and $\sigma > 0$.

Since the principal's expected long-run per-period utility for any $m > 0$ is bounded from above by the corresponding quantity under the reward policy in which m is optimized, we have the desired result. This completes our proof of Theorem 5. ■

Appendix C: Proofs of Results in Section 7

In this appendix, we first formally define the game and the equilibrium concept for the multi-agent budget-constrained setting defined in Section 7. Subsequently, we present the proofs of our results in Section 7.

C.1. Formal Definition of the Game and the Equilibrium Concept

The principal announces a common reward policy that specifies, for each period, a reward to each agent as a function of the past performance of the agents. Formally, define a *reward policy* π as a sequence of functions $\{\pi_{n,t};\ n = 1, 2, \ldots, N, \text{ and } t \geq 1\}$, where, for any (n,t), the function $\pi_{n,t} : \mathbb{R}_+^{N\cdot(t-1)} \to \{0,1\}$ maps the history of outputs of all agents in the first $t-1$ periods (denoted by $\mathbf{q}_{t-1}$) to 1 if the principal rewards agent n in period t, and 0 otherwise. We let Π denote the set of all such reward policies.

Given a reward policy $\pi \in \Pi$, the agents optimally decide their effort policies in equilibrium. Formally, for any $n = 1, 2, \ldots, N$, an *effort policy* λ_n for agent n is defined by a sequence of functions $\{\lambda_{n,t};\ t \geq 1\}$, where, for any t, the function $\lambda_{n,t} : \mathbb{R}_+^{N\cdot(t-1)} \to \mathbb{R}_+$ maps the history of outputs $\mathbf{q}_{t-1}$ of all agents in the first $(t-1)$ periods to an effort level for agent n in period t. Let Λ denote the set of all possible effort policies of the agents. For any n, let $\boldsymbol{\lambda}_{-n} = (\lambda_1, \ldots, \lambda_{n-1}, \lambda_{n+1}, \ldots, \lambda_N)$ denote the vector of the effort policies adopted by all agents except agent n. Let

$$U_n(\lambda_n, \boldsymbol{\lambda}_{-n} \mid \pi) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}(v \cdot \pi_{n,t} - c \cdot \lambda_{n,t})\right]$$

denote agent n's expected long-run per-period utility under the reward policy π. For simplicity of exposition, we have suppressed the dependence of $\pi_{n,t}$ and $\lambda_{n,t}$ on the history of past outputs $\mathbf{q}_{t-1}$, which, in turn, is a function of random variables $W_{m,s}$, $m = 1, 2, \ldots, N$ and $s = 1, 2, \ldots, t-1$. Further, the expectation in the above expression for $U_n(\cdot)$ is taken with respect to the outputs of all agents.

We use the following equilibrium concept:

Definition C.1. Given a reward policy π, an effort policy $\lambda \in \Lambda$ constitutes a Nash equilibrium if, for each n, assuming that each agent $m \neq n$ exerts effort according to the policy $\lambda_m := \lambda$, agent n's expected long-run per-period utility is maximized at the effort policy λ. Precisely,

$$U_n(\lambda, \boldsymbol{\lambda}_{-n} \mid \pi) \ \geq \ U_n(\widehat{\lambda}, \boldsymbol{\lambda}_{-n} \mid \pi) \quad \forall \widehat{\lambda} \in \Lambda, n.$$

The principal's problem in the multi-agent, budget-constrained setting of Section 7 can be formally stated as follows:

$$\begin{aligned}
\max_{\substack{\pi\in\Pi \\ \lambda\in\Lambda}} \quad & \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{n=1}^{N}\mathbb{E}\Big[v_0 \cdot W_{n,t}\cdot\lambda_t \ - \ c_0\cdot\pi_{n,t}\Big] \\
\text{s.t.} \quad & U_n(\lambda, \boldsymbol{\lambda}_{-n} \mid \pi) \ \geq \ U_n(\widehat{\lambda}, \boldsymbol{\lambda}_{-n} \mid \pi) \quad \forall \widehat{\lambda} \in \Lambda, n, \\
& U_n(\lambda, \boldsymbol{\lambda}_{-n} \mid \pi) \ \geq \ 0 \quad \forall n, \\
& \sum_{n=1}^{N} \pi_{n,t} \leq B \quad \forall t.
\end{aligned}$$

C.2. Proofs of Results in Section 7

Proof of Lemma 9. To prove this result, we first establish two intermediate results, namely Lemmas C.1 and C.2.

Lemma C.1. *Consider any agent, say n, and suppose that the agent uses the effort policy $\lambda_{n,t} = B/(N\theta)$ for all t if $r \geq 1/\theta$ and $\lambda_{n,t} = 0$ for all t if $r < 1/\theta$. Then, irrespective of the actions of the other agents, under the modified score-based policy, we have*

$$\lim_{T\to\infty}\frac{\mathbb{E}[S_{n,T+1}]}{T} = 0.$$

Proof of Lemma C.1. First, consider the case $r < 1/\theta$. Since $\lambda_{n,t} = 0$ for all t and $S_{n,1} = 0$, we have $\pi^{SB}_{n,t} = 0$ and $S_{n,t} = 0$ for all t. Thus, the desired result holds. For the remainder of the proof, we focus on $r \geq 1/\theta$ in which case $\lambda_{n,t} = B/(N\theta)$ for all t. Define a sequence $\{Z_t, t \in \mathbb{N}\}$ of independent and identically distributed random variables, where, for any t,

$$Z_t = \begin{cases} 1, & \text{with probability } B/N \\ 0, & \text{with probability } 1 - B/N. \end{cases} \tag{C.1}$$

Define a sequence $\{X_{n,t}, t \in \mathbb{N}\}$ as follows:

$$X_{n,t+1} = X_{n,t} - \min(X_{n,t}, Z_t) + \theta W_{n,t}\lambda_{n,t}, \tag{C.2}$$

where $X_{n,1} = 0$. Using the identity $a + b = \min(a,b) + \max(a,b)$ in (C.2), we have $X_{n,t+1} = (X_{n,t} - Z_t)^+ + \theta W_{n,t}\lambda_{n,t}$. Since $W_{n,t} \in \mathbb{R}_+$ and $\lambda_{n,t} \in \mathbb{R}_+$, we have $X_{n,t} \geq 0$ for all $t \in \mathbb{N}$. Define another sequence $\{Y_{n,t}, t \in \mathbb{N}\}$ where $Y_{n,t} = (X_{n,t} - Z_t)^+$. Then, we have:

$$Y_{n,t+1} = (X_{n,t+1} - Z_{t+1})^+ = (Y_{n,t} + \theta W_{n,t}\lambda_{n,t} - Z_{t+1})^+.$$

Using $\lambda_{n,t} = B/(N\theta)$ and $\mathbb{E}[W_{n,t}] = 1$ for all t, we have $\mathbb{E}[\theta W_{n,t}\lambda_{n,t}] = B/N = \mathbb{E}[Z_{t+1}]$. Then, from Theorem 4 of Loynes (1962), we have:

$$\lim_{T\to\infty} \frac{\mathbb{E}[Y_{n,T}]}{T} = 0. \tag{C.3}$$

Thus, we have:

$$\begin{aligned} \lim_{T\to\infty} \frac{\mathbb{E}[X_{n,T+1}]}{T} &= \lim_{T\to\infty} \frac{\mathbb{E}[(X_{n,T} - Z_T)^+ + \theta W_{n,T}\lambda_{n,T}]}{T} \\ &= \lim_{T\to\infty} \frac{\mathbb{E}[Y_{n,T} + \theta W_{n,T}\lambda_{n,T}]}{T} \\ &= \underbrace{\lim_{T\to\infty} \frac{\mathbb{E}[Y_{n,T}]}{T}}_{=0 \quad ((\text{C.3}))} + \underbrace{\lim_{T\to\infty} \frac{\mathbb{E}[\theta W_{n,T}\lambda_{n,T}]}{T}}_{=0 \quad (\because \mathbb{E}[W_{n,T}]=1;\, \lambda_{n,T}=B/(N\theta))} \\ &= 0. \end{aligned} \tag{C.4}$$

For any $p \in [B/N, 1]$, define a sequence $\{\hat{Z}_t(p), t \in \mathbb{N}\}$ of random variables where, for any t,

$$\hat{Z}_t(p) = \begin{cases} 1, & \text{with probability } \frac{p - B/N}{1 - B/N} \\ 0, & \text{otherwise.} \end{cases} \tag{C.5}$$

We show

$$S_{n,t} \leq_{st} X_{n,t} + 1 \quad \forall t \in \mathbb{N}. \tag{C.6}$$

We use induction to prove (C.6). Since $S_{n,1} = X_{n,1} = 0$, (C.6) holds for $t = 1$. Suppose, for the induction hypothesis, that (C.6) holds for some $t \geq 1$. We consider the following cases:

(i) $\underline{S_{n,t} < 1}$. Then, under the modified score-based policy, we have $\pi^{SB}_{n,t} = 0$. Also,

$$X_{n,t+1} = X_{n,t} - \min(X_{n,t}, Z_t) + \theta W_{n,t}\lambda_{n,t} \geq \theta W_{n,t}\lambda_{n,t}.$$

Further, we have

$$S_{n,t+1} = S_{n,t} - \pi^{SB}_{n,t} + \theta W_{n,t}\lambda_{n,t} < 1 + \theta W_{n,t}\lambda_{n,t} \leq 1 + X_{n,t+1}.$$

(ii) $\underline{S_{n,t} \geq 1}$. Under the modified score-based policy, suppose that there are $M \geq 1$ agents (including agent n) whose scores in period t are at least 1. There are two possibilities: (a) If $M \leq B$, then agent n is rewarded in t with probability 1. (b) If $M > B$, then according to the tie-breaking rule in the modified score-based policy, agent n is rewarded in t with probability B/M. Combining (a) and (b), the probability that agent n is rewarded in period t given $S_{n,t} \geq 1$ is $\min\{1, B/M\}$. Then, the following holds:

$$\pi_{n,t}^{SB} \;=_{st}\; Z_t + (1 - Z_t) \cdot \hat{Z}_t(p_{n,t}), \quad \text{where} \quad p_{n,t} = \min\{1, B/M\}. \tag{C.7}$$

The above relationship deserves some explanation: Using the definition of Z_t and $\hat{Z}_t(\cdot)$ from (C.1) and (C.5), and the fact that Z_t and $\hat{Z}_t(\cdot)$ are independent, by construction, $Z_t + (1 - Z_t) \cdot \hat{Z}_t(p_{n,t})$ is binary and takes the value equal to 1 with probability $p_{n,t} = \min\{1, B/M\}$ and 0 otherwise, where $p_{n,t} \geq B/N$ since $M \leq N$ and $B < N$. This is consistent with the probabilistic tie breaking that agent n experiences in period t given $S_{n,t} \geq 1$, and establishes (C.7). Further, since $(1 - Z_t) \cdot \hat{Z}_t(p_{n,t})$ is non-negative with probability 1, using Theorem 1.A.1 in Shaked and Shanthikumar (2007), we have $\pi_{n,t}^{SB} \geq_{st} Z_t$.

Thus, we have:

$$\begin{aligned}
S_{n,t+1} &= S_{n,t} - \pi_{n,t}^{SB} + \theta W_{n,t} \lambda_{n,t} \\
&\leq_{st} 1 + X_{n,t} - \pi_{n,t}^{SB} + \theta W_{n,t} \lambda_{n,t} \quad \text{(using induction hypothesis)} \\
&\leq_{st} 1 + X_{n,t} - Z_t + \theta W_{n,t} \lambda_{n,t} \quad \text{(using } \pi_{n,t}^{SB} \geq_{st} Z_t\text{)} \\
&\leq_{st} 1 + X_{n,t} - \min(X_{n,t}, Z_t) + \theta W_{n,t} \lambda_{n,t} \\
&= 1 + X_{n,t+1}.
\end{aligned}$$

Thus, from the above two cases, we conclude the induction argument and hence the proof of (C.6).

Using (C.6), we have:

$$\begin{aligned}
& \lim_{T\to\infty} \frac{\mathbb{E}[S_{n,T+1}]}{T} \;\leq\; \lim_{T\to\infty} \frac{1 + \mathbb{E}[X_{n,T+1}]}{T} \\
\implies \quad & \lim_{T\to\infty} \frac{\mathbb{E}[S_{n,T+1}]}{T} \;\leq\; 0 \qquad \text{(using (C.4))} \\
\implies \quad & \lim_{T\to\infty} \frac{\mathbb{E}[S_{n,T+1}]}{T} = 0, \quad \text{(since } S_{n,t} \geq 0 \ \forall t\text{)},
\end{aligned}$$

thus achieving the desired result and completing the proof of Lemma C.1. □

Remark: The proof of Lemma C.1 assumes that the vector of starting scores of the agents $\mathbf{S}_1 = (S_{1,1}, S_{2,1}, \ldots, S_{N,1}) = \mathbf{0}$. The proof easily extends for any arbitrary vector of starting scores of the agents. We highlight the main changes. Consider any agent, say n. Since the scores are always non-negative, we let $S_{n,1} = s$ for an arbitrary and finite $s \geq 0$. Consider the case $r < 1/\theta$. Let $\overline{\lambda}_n := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\lambda_{n,t}]$. Under the modified score-based reward policy with parameter θ, we have:

$$\begin{aligned}
& \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\pi_{n,t}^{SB}] = \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[S_{n,t} - S_{n,t+1} + \theta W_{n,t} \lambda_{n,t}\right] \\
&= \lim_{T\to\infty} \frac{\mathbb{E}[S_{n,1}]}{T} - \lim_{T\to\infty} \frac{\mathbb{E}[S_{n,T+1}]}{T} + \theta \overline{\lambda}_n \quad \text{(using the independence of } W_{n,t} \text{ across } t \text{ and } \mathbb{E}[W_{n,t}] = 1 \ \forall t\text{)} \\
&= - \lim_{T\to\infty} \frac{\mathbb{E}[S_{n,T+1}]}{T} + \theta \overline{\lambda}_n \quad \text{(since } S_{n,1} = s \text{ and } s \text{ is finite).}
\end{aligned}$$

Since $\lambda_{n,t}=0$ for all t, agent n's score either remains the same over time or decreases as a result of the rewards received by the principal. Further, since $S_{n,1}=s$ is finite and $\lambda_{n,t}=0$ for all t, the agent is rewarded for only a finite number of periods, and thus, we have $\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\pi_{n,t}^{SB}]=0$. Thus, $\lim_{T\to\infty}\frac{\mathbb{E}[S_{n,T+1}]}{T}=0$.

For the case $r\geq 1/\theta$, the only change that is required in the proof is to initialize the sequence $\{X_{n,t},t\in\mathbb{N}\}$ as $X_{n,1}=s$. The remainder of the proof steps remain the same as before. □

LEMMA C.2. *Consider an arbitrary $n=1,2,\ldots,N$. Under the modified score-based reward policy with parameter θ, assume that every agent $m\neq n$ follows the equilibrium effort policy, i.e., $\lambda_{m,t}=B/(N\theta)$ for all t if $r\geq 1/\theta$ and $\lambda_{m,t}=0$ for all t if $r<1/\theta$. The expected long-run per-period utility of agent n using an arbitrary effort policy λ is bounded from above as follows: (i) If $r\geq 1/\theta$, then the upper bound is $(B/N)\cdot(v-c/\theta)$. (ii) If $r<1/\theta$, then the upper bound is 0.*

Proof of Lemma C.2. Let $\overline{\lambda}_m:=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\lambda_{m,t}]$. Under the modified score-based reward policy with parameter θ, we have:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\pi_{m,t}^{SB}]=\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[S_{m,t}-S_{m,t+1}+\theta W_{m,t}\lambda_{m,t}\right]=-\lim_{T\to\infty}\frac{\mathbb{E}[S_{m,T+1}]}{T}+\theta\overline{\lambda}_m. \tag{C.8}$$

Consider the case $r\geq 1/\theta$. Consider any agent $m\neq n$. Using $\lambda_{m,t}=B/(N\theta)$ for all t, (C.8) and Lemma C.1, we have

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\pi_{m,t}^{SB}]=-\lim_{T\to\infty}\frac{\mathbb{E}[S_{m,T+1}]}{T}+\theta\overline{\lambda}_m=\frac{B}{N}\quad\forall m\neq n. \tag{C.9}$$

Also, for any t, we know that $\sum_{m=1}^{N}\pi_{m,t}^{SB}\leq B$. This implies the following:

$$\begin{aligned}&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\pi_{n,t}^{SB}\leq B-\sum_{m\neq n}\left(\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\pi_{m,t}^{SB}\right)\\ \Longrightarrow\quad&\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\pi_{n,t}^{SB}\leq\frac{B}{N}\quad\text{(using (C.9))}.\end{aligned}$$

Using the above relationship, together with $S_{n,t}\geq 0$ in (C.8), we have:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[v\cdot\pi_{n,t}^{SB}-c\cdot\lambda_{n,t}\right]\leq v\cdot\min(\theta\overline{\lambda}_n,B/N)-c\overline{\lambda}_n. \tag{C.10}$$

Thus, an upper bound on agent n's expected long-run per-period utility can be obtained by maximizing (C.10) with respect to $\overline{\lambda}_n$. Note that any choice of $\overline{\lambda}_n>B/(N\theta)$ is dominated by $\overline{\lambda}_n=B/(N\theta)$. Therefore, it is sufficient to consider:

$$\max_{0\leq\overline{\lambda}_n\leq B/(N\theta)}\quad\overline{\lambda}_n\cdot(v\theta-c)\,.$$

Since $v/c\geq 1/\theta$, the optimal solution to the above optimization problem is $\overline{\lambda}_n=B/(N\theta)$. Thus, the desired upper bound in this case is $(B/N)\cdot(v-c/\theta)$.

Consider the case $r<1/\theta$. The proof for this case follows the same logic as in the previous case except for a few differences. For brevity, we only highlight the differences. Since agent n's expected long-run per-period reward is at most 1, using (C.8), we have:

$$\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}\left[v\cdot\pi_{n,t}^{SB}-c\cdot\lambda_{n,t}\right]\leq v\cdot\min(\theta\overline{\lambda}_n,1)-c\overline{\lambda}_n.$$

Again, maximizing the above expression with respect to $\overline{\lambda}_n$ and noting that the objective function is decreasing in $\overline{\lambda}_n > 1/\theta$, we have:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[v\cdot \pi^{SB}_{n,t} - c\cdot \lambda_{n,t}\right] \le \max_{0\le \overline{\lambda}_n \le 1/\theta} \overline{\lambda}_n \cdot (v\theta - c).$$

Since $v/c < 1/\theta$, the maximum of the above problem is obtained at $\overline{\lambda}_n = 0$ and thus the desired upper bound is 0. This completes our proof of Lemma C.2. □

We are now ready to prove Lemma 9. Under the modified score-based reward policy with parameter θ, suppose that every agent $m \ne n$ exerts the effort $\lambda_{m,t} = B/(N\theta)$ for all t if $r \ge 1/\theta$, and $\lambda_{m,t} = 0$ for all t if $r < 1/\theta$. We show that agent n cannot benefit by deviating from the equilibrium. We demonstrate this result by showing that under the equilibrium effort policy, the upper bound on agent n's expected long-run per-period utility characterized in Lemma C.2 is tight.

Suppose that agent n follows the equilibrium effort policy, i.e., $\lambda_{n,t} = B/(N\theta)$ for all t if $r \ge 1/\theta$, and $\lambda_{n,t} = 0$ for all t if $r < 1/\theta$. Then, we have:

$$\begin{aligned}
&\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[v\cdot \pi^{SB}_{n,t} - c\cdot \lambda_{n,t}\right] = v\cdot\left(-\lim_{T\to\infty}\frac{\mathbb{E}[S_{n,T+1}]}{T} + \theta\overline{\lambda}_n\right) - c\overline{\lambda}_n \quad \text{(using (C.8))}\\
&= \overline{\lambda}_n \cdot (v\theta - c) \quad \text{(using Lemma C.1)}\\
&= \begin{cases} (B/N)\cdot(v - c/\theta) & \text{if} \quad r \ge 1/\theta\\ 0 & \text{if} \quad r < 1/\theta. \end{cases}
\end{aligned}$$

The above relationship, together with Lemma C.2, establish the desired result. ■

Proof of Theorem 7. To prove this result, we first establish an intermediate result below.

LEMMA C.3. *For the multi-agent budget-constrained extension in Section 7, for any reward policy, the principal's expected long-run per-period utility is upper bounded by* $(v_0 r - c_0)\cdot B$.

Proof of Lemma C.3. Consider any reward policy π. In response to the reward policy, suppose that each agent n uses an effort policy λ^*_n, in equilibrium. Let $\overline{\pi}_n$ and $\overline{\lambda}^*_n$ denote the expected long-run per-period values corresponding to $\pi_{n,t}$ and $\lambda^*_{n,t}$. That is,

$$\overline{\pi}_n = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\pi_{n,t}] \quad \text{and} \quad \overline{\lambda}^*_n = \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\lambda^*_{n,t}].$$

Since the agents' expected long-run per-period utilities are non-negative in equilibrium, we have $v\overline{\pi}_n - c\overline{\lambda}^*_n \ge 0$ for all n. This holds because an agent could always exert zero effort in every period.

The principal's expected long-run per-period utility under reward policy π is given as:

$$\sum_{n=1}^{N}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[v_0 W_{n,t}\lambda^*_{n,t} - c_0\pi_{n,t}] = \sum_{n=1}^{N}(v_0\overline{\lambda}^*_n - c_0\overline{\pi}_n) \le \sum_{n=1}^{N}(v_0 r - c_0)\overline{\pi}_n, \tag{C.11}$$

where we have used the inequality $v\overline{\pi}_n - c\overline{\lambda}^*_n \ge 0$ for all n.

Further, we know that $\sum_{n=1}^{N}\pi_{n,t} \le B$ for all t. Thus, $\sum_{n=1}^{N}\overline{\pi}_n \le B$. Using this and the assumption that $v_0 r - c_0 \ge 0$ in (C.11), we have:

$$\sum_{n=1}^{N}\lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[v_0 W_{n,t}\lambda^*_{n,t} - c_0\pi_{n,t}] \le (v_0 r - c_0)B,$$

completing the proof of Lemma C.3. □

We are now ready to prove Theorem 7. From Lemma 9, under the modified score-based policy with parameter θ, in equilibrium, every agent exerts a per-period effort equal to $B/(N\theta)$ if $r \geq 1/\theta$, and equal to 0 otherwise. Further, using Lemma C.1, in equilibrium, for any n, we have

$$\lim_{T\to\infty} \frac{\mathbb{E}[S_{n,T+1}]}{T} = 0.$$

Using (C.8), for any n, we thus have:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\pi_{n,t}^{SB}] = -\lim_{T\to\infty} \frac{\mathbb{E}[S_{n,T+1}]}{T} + \theta\overline{\lambda}_n = \theta\overline{\lambda}_n. \tag{C.12}$$

The principal's expected long-run per-period utility under modified score-based reward policy is given as:

$$\begin{aligned}\sum_{n=1}^{N} \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[v_0 W_{n,t}\lambda_{n,t} - c_0\pi_{n,t}^{SB}] &= \sum_{n=1}^{N} \mathbb{E}[v_0\overline{\lambda}_n - c_0\theta\overline{\lambda}_n] \quad \text{(using (C.12))}\\ &= \begin{cases}(v_0/\theta - c_0)\cdot B & \text{if } 1/\theta \leq r\\ 0 & \text{if } 1/\theta > r.\end{cases}\end{aligned}$$

Thus, the principal's expected utility is maximized by setting $1/\theta = r$ and is equal to the value $(v_0 r - c_0)\cdot B \geq 0$, which is equal to the upper bound obtained in Lemma C.3. This completes the proof of Theorem 7. ■

Appendix D: Multi-Agent, Budget-Constrained and Private-Value Setting

In this appendix, we address Remark 5 of Section 7. This appendix is organized as follows: First, we formally specify the multi-agent, budget-constrained and private-value setting, and formulate the principal's problem under this setting. Second, we state our main result, i.e., Theorem D.1 for this setting. Finally, we present the proof of Theorem D.1.

Problem Setting and Formulation: There are a total of N agents. The principal can offer a maximum of B rewards to the agents in any given period, where $B := \widehat{B}\cdot N$ with $\widehat{B} \in (0,1)$. Agent n's value-to-cost ratio is $r_n := v_n/c_n$, where r_n is privately known to agent n and is a realization of a random variable R_n with cumulative distribution function F (and probability density function f) over the support $[0,\infty)$. We assume that R_n is independently and identically distributed across n. Further, R_n is independent of $W_{n,t}$, for all (n,t). Let $\overline{F}(\cdot) = 1 - F(\cdot)$. Recall that

$$\psi(r) \;:=\; r - \left(\frac{1-F(r)}{f(r)}\right), \tag{16}$$

and $\psi(r)$ is strictly increasing in r. Further, recall that $\psi^{-1}(\cdot)$ denote the inverse of the function $\psi(\cdot)$. All other details are the same as those in the multi-agent budget-constrained setting analyzed in Section 7 (where the value-to-cost ratio of the agents is common and publicly-known). The principal's goal is to obtain a reward policy that maximizes her expected utility per-period over the long-run. For simplicity in exposition, we reuse our notation of the base model, and denote the principal's problem under this extension also by $\mathcal{P}$ and its optimal value by OPT.

We now express the principal's problem formally. As in Appendix C, given a reward policy $\pi \in \Pi$, the agents optimally decide their effort policies in equilibrium. Let Λ denote the set of all possible effort policies of the agents. Further, for any n, let λ_n denote the effort policy of agent n and let $\boldsymbol{\lambda}_{-n} = (\lambda_1, \ldots, \lambda_{n-1}, \lambda_{n+1}, \ldots, \lambda_N)$ denote the vector of the effort policies of all agents excluding agent n. Also, let

$$U_n(\lambda_n, \boldsymbol{\lambda}_{-n} \mid \pi, c_n, v_n) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}(v_n \cdot \pi_{n,t} - c_n \cdot \lambda_{n,t})\right]$$

denote agent n's expected long-run per-period utility under the reward policy π. For notational simplicity, we have suppressed the dependence of $\pi_{n,t}$ on the history of past outputs of the agents, and of $\lambda_{n,t}$ on the history of past outputs of the agents and agent n's attribute (c_n, v_n). Also, in the above expression for $U_n(\cdot)$, the expectation is taken with respect to the outputs of all agents and the value-to-cost ratios of all agents except agent n.

We use the following equilibrium concept:

Definition D.1. Given a reward policy π, an effort policy $\lambda(c, v) \in \Lambda$ constitutes a Bayesian Nash equilibrium if, for each n, assuming that each agent $m \neq n$ exerts effort according to the policy $\lambda_m := \lambda(c_m, v_m)$, agent n's expected long-run per-period utility is maximized at the effort policy $\lambda_n := \lambda(c_n, v_n)$. Precisely,

$$U_n(\lambda_n, \boldsymbol{\lambda}_{-n} \mid \pi, c_n, v_n) \;\geq\; U_n(\widehat{\lambda}, \boldsymbol{\lambda}_{-n} \mid \pi, c_n, v_n) \quad \forall \widehat{\lambda} \in \Lambda, c_n, v_n, n. \tag{D.1}$$

The principal's problem under the multi-agent, budget-constrained, and private-value setting is formulated as follows:

$$\begin{aligned} OPT \;:=\; \max_{\substack{\pi\in\Pi \\ \lambda(\cdot)\in\Lambda}} \quad & \lim_{T\to\infty}\frac{1}{T}\sum_{t=1}^{T}\sum_{n=1}^{N}\mathbb{E}\Big[v_0 \cdot W_{n,t}\cdot \lambda_{n,t} \;-\; c_0 \cdot \pi_{n,t}\Big] \\ \text{s.t.} \quad & U_n(\lambda_n, \boldsymbol{\lambda}_{-n} \mid \pi, c_n, v_n) \;\geq\; U_n(\widehat{\lambda}, \boldsymbol{\lambda}_{-n} \mid \pi, c_n, v_n) \quad \forall \widehat{\lambda} \in \Lambda, c_n, v_n, n, & \text{(D.1)} \\ & U_n(\lambda_n, \boldsymbol{\lambda}_{-n} \mid \pi, c_n, v_n) \;\geq\; 0 \quad \forall c_n, v_n, n, \\ & \sum_{n=1}^{N} \pi_{n,t} \leq B \quad \forall t. & \text{(D.2)} \end{aligned}$$

To state our main result for the multi-agent, budget-constrained, and private-value setting, we need the following definition.

Definition D.2. Given a reward policy π and a fixed $\epsilon > 0$, an effort policy $\lambda(c, v) \in \Lambda$ constitutes an ϵ-Bayesian Nash equilibrium if, for each n, assuming that each agent $m \neq n$ exerts effort according to the policy $\lambda_m := \lambda(c_m, v_m)$, the loss in agent n's expected long-run per-period utility from using the effort policy $\lambda_n := \lambda(c_n, v_n)$ relative to any other effort policy is at most ϵ. Precisely,

$$U_n(\lambda_n, \boldsymbol{\lambda}_{-n} \mid \pi, c_n, v_n) \;\geq\; U_n(\widehat{\lambda}, \boldsymbol{\lambda}_{-n} \mid \pi, c_n, v_n) \;-\; \epsilon \quad \forall \widehat{\lambda} \in \Lambda, c_n, v_n, n. \tag{D.1'}$$

We note that inequality (D.1') is an ϵ-relaxation of inequality (D.1). Our main result for the multi-agent, budget-constrained, and private-value setting is as follows:

Theorem D.1. *Given any $\epsilon > 0$, there exists $N(\epsilon) \geq 1$ such that for all $N \geq N(\epsilon)$, under the modified score-based reward policy with the parameter θ defined by*

$$\frac{1}{\theta} = \begin{cases} \overline{F}^{-1}(\widehat{B}) & \text{if } \widehat{B} \leq \overline{F}\left(\psi^{-1}(c_0/v_0)\right), \\ \psi^{-1}(c_0/v_0) & \text{otherwise,} \end{cases} \tag{D.3}$$

the effort policy

$$\lambda(c,v)=\begin{cases}1/\theta, & \textit{if } v/c\geq 1/\theta\\ 0, & \textit{otherwise}\end{cases}\tag{D.4}$$

forms an ϵ-Bayesian Nash equilibrium, and the principal's expected per-period utility under this reward policy is at least OPT.

D.1. Proof of Theorem D.1

This proof is organized as follows: First, we obtain an upper bound on the principal's optimal value OPT in Appendix D.1.1 (see Lemma D.1). Next, we analyze the class of modified score-based policies for the multi-agent, budget-constrained, and private-value setting in Appendix D.1.2. This analysis involves establishing several intermediate results, namely Lemmas D.2–D.8. After proving these lemmas, we will complete our proof of Theorem D.1 in Appendix D.1.3.

D.1.1. An Upper Bound on OPT.

Lemma D.1. *An upper bound on OPT is obtained as follows:*

$$OPT\leq\overline{OPT}:=N\cdot\left(v_0\cdot\psi^{-1}\left(\frac{c_0+\rho^*}{v_0}\right)-c_0\right)\cdot\left[1-F\left(\psi^{-1}\left(\frac{c_0+\rho^*}{v_0}\right)\right)\right],$$

where

$$\rho^*=\begin{cases}v_0\psi\left(\overline{F}^{-1}(\widehat{B})\right)-c_0 & \textit{if } \widehat{B}\leq\overline{F}\left(\psi^{-1}(c_0/v_0)\right),\\ 0 & \textit{otherwise.}\end{cases}\tag{D.5}$$

Proof of Lemma D.1. We obtain an upper bound on OPT by relaxing the moral-hazard friction faced by the principal (i.e., assuming that the agents' efforts are observable to the principal), and by relaxing the budget constraint specified in (D.2) to hold in expectation (with respect to the uncertainties underlying the agents' outputs, and their privately-known value-to-cost ratios) and over the long-run. We denote the relaxed problem as $\overline{\mathcal{P}}$ and its optimal value as $\overline{OPT}$. For better exposition, the formulation of $\overline{\mathcal{P}}$ is outlined in Appendix E. As shown in Appendix E, problem $\overline{\mathcal{P}}$ can be formally stated as follows:

$$\overline{OPT}\;=\;\max_{\{\overline{\pi}_n(\cdot),\overline{\lambda}_n(\cdot)\}\ \forall n}\quad\sum_{n=1}^{N}\mathbb{E}\left[v_0\cdot\psi(R_n)\overline{\pi}_n(R_n)-v_0y_n(0)-c_0\cdot\overline{\pi}_n(R_n)\right]\quad\text{s.t. (E.7)–(E.10).}$$

Define $\{\overline{\pi}^*_n(\cdot),\overline{\lambda}^*_n(\cdot)\}$ as follows:

$$\overline{\pi}^*_n(r_n)=\begin{cases}1 & \psi(r_n)\geq(c_0+\rho^*)/v_0,\\ 0 & \text{otherwise,}\end{cases}\quad\text{and}\quad\overline{\lambda}^*_n(r_n)=\begin{cases}\psi^{-1}((c_0+\rho^*)/v_0) & \psi(r_n)\geq(c_0+\rho^*)/v_0,\\ 0 & \text{otherwise,}\end{cases}\tag{D.6}$$

where ρ^* is specified in (D.5). We will show below that the above solution is optimal for problem $\overline{\mathcal{P}}$.

Since $y_n(0)\geq 0$ for all n (see (E.10)), we have:

$$\overline{OPT}\;\leq\;\widehat{OPT}\;:=\;\max_{\overline{\pi}_n(\cdot)\ \forall n}\quad\mathbb{E}\left[\sum_{n=1}^{N}[v_0\psi(R_n)-c_0]\cdot\overline{\pi}_n(R_n)\right]\quad\text{s.t. (E.7),}$$

where we recall that the constraint (E.7) is

$$\sum_{n=1}^{N}\mathbb{E}[\overline{\pi}_n(R_n)]\leq B.\tag{E.7}$$

We obtain an optimal solution for the above problem (whose optimal value is $\widehat{OPT}$) by solving its Lagrangian relaxation. That is, we identify $\{\overline{\pi}_n(\cdot); n = 1, 2, \ldots, N\}$ and ρ (which represents the Lagrange multiplier corresponding to (E.7)) that simultaneously satisfy the following conditions:

$$\{\overline{\pi}_n(\cdot)\} = \underset{\widehat{\pi}_n(\cdot)\in[0,1]\ \forall n}{\arg\max}\ \mathbb{E}\left[\sum_{n=1}^{N}[v_0\psi(R_n) - c_0]\cdot\widehat{\pi}_n(R_n) + \rho\cdot\left(B - \sum_{n=1}^{N}\widehat{\pi}_n(R_n)\right)\right], \tag{D.7}$$

$$\rho\cdot\left(B - \sum_{n=1}^{N}\mathbb{E}[\overline{\pi}_n(R_n)]\right) = 0, \tag{D.8}$$

$$\rho \geq 0, \quad \text{and} \quad \sum_{n=1}^{N}\mathbb{E}[\overline{\pi}_n(R_n)] \leq B. \tag{D.9}$$

In our steps below, we show that $\overline{\pi}^*_n(\cdot)$ in (D.6) and ρ^* in (D.5), together satisfy the conditions (D.7)–(D.9), and subsequently, compute $\widehat{OPT}$.

First, note that given $\rho = \rho^*$, $\overline{\pi}^*_n(\cdot)$ satisfies (D.7). Consider the case $\underline{\widehat{B} \leq \overline{F}\left(\psi^{-1}(c_0/v_0)\right)}$. It is easy to check that ρ^* in (D.5) satisfies $\widehat{B} = \overline{F}\left(\psi^{-1}((c_0+\rho^*)/v_0)\right)$. Further, since $\psi^{-1}(\cdot)$ is an increasing function, where $\psi(\cdot)$ is defined in (16), $\rho^* \geq 0$. Using $\overline{\pi}^*_n(\cdot)$ from (D.6), we have:

$$B - \mathbb{E}\left[\sum_{n=1}^{N}\overline{\pi}^*_n(R_n)\right] = N\left[\widehat{B} - \overline{F}\left(\psi^{-1}((c_0+\rho^*)/v_0)\right)\right] = 0,$$

and thus, (D.7)–(D.9) are satisfied.

Consider the case $\underline{\widehat{B} > \overline{F}\left(\psi^{-1}(c_0/v_0)\right)}$. From (D.5), $\rho^* = 0$. Using $\overline{\pi}^*_n(\cdot)$ from (D.6), we have:

$$B - \mathbb{E}\left[\sum_{n=1}^{N}\overline{\pi}^*_n(R_n)\right] = N\left[\widehat{B} - \overline{F}\left(\psi^{-1}((c_0+\rho^*)/v_0)\right)\right] = N\left[\widehat{B} - \overline{F}\left(\psi^{-1}(c_0/v_0)\right)\right] > 0,$$

and thus, (D.7)–(D.9) are satisfied.

Thus, we have:

$$\widehat{OPT} = \mathbb{E}\left[\sum_{n=1}^{N}[v_0\psi(R_n) - c_0]\cdot\overline{\pi}^*_n(R_n)\right]. \tag{D.10}$$

So far, we have obtained an upper bound on $\overline{OPT}$. Next, we show that the solution $\{\overline{\pi}^*_n(\cdot), \overline{\lambda}^*_n(\cdot)\}$ in (D.6) is feasible to $\overline{\mathcal{P}}$ and achieves the upper bound $\widehat{OPT}$ on $\overline{OPT}$.

As we have already shown in the two cases above, the given solution satisfies (E.7). Using the assumption that $\psi(\cdot)$ is strictly increasing and the fact that $\psi(0) = -1/f(0) < 0$, we have $\psi^{-1}((c_0+\rho^*)/v_0) \geq \psi^{-1}(0) > 0$. Thus, from (D.6), $y^*_n(0) = 0\cdot\overline{\pi}^*_n(0) - \overline{\lambda}^*_n(0) = 0$; thus, the constraint (E.10) holds at equality. Further, $\overline{\pi}^*_n(r)$ is non-decreasing in r, and thus it satisfies (E.9). Also, it is easy to verify that $\{\overline{\pi}^*_n(\cdot), \overline{\lambda}^*_n(\cdot)\}$ satisfies (E.8). Thus, $\{\overline{\pi}^*_n(\cdot), \overline{\lambda}^*_n(\cdot)\}$ is feasible to $\overline{\mathcal{P}}$. Further, it achieves the upper bound on $\overline{OPT}$, i.e.,

$$\sum_{n=1}^{N}\mathbb{E}[v_0\cdot\psi(R_n)\overline{\pi}^*_n(R_n) - v_0 y^*_n(0) - c_0\cdot\overline{\pi}^*_n(R_n)] = \sum_{n=1}^{N}\mathbb{E}[v_0\cdot\psi(R_n)\overline{\pi}^*_n(R_n) - c_0\cdot\overline{\pi}^*_n(R_n)] = \widehat{OPT},$$

proving the optimality of $\{\overline{\pi}^*_n(\cdot), \overline{\lambda}^*_n(\cdot)\}$ for $\overline{\mathcal{P}}$.

We now express $\overline{OPT}$ in closed-form.

$$\begin{aligned}\overline{OPT} &= \sum_{n=1}^{N}\mathbb{E}[v_0\overline{\lambda}^*_n(R_n) - c_0\overline{\pi}^*_n(R_n)] = \sum_{n=1}^{N}\left(v_0\cdot\psi^{-1}((c_0+\rho^*)/v_0) - c_0\right)\cdot\mathbb{P}\left[R_n \geq \psi^{-1}((c_0+\rho^*)/v_0)\right] \\ &= N\cdot\left(v_0\cdot\psi^{-1}((c_0+\rho^*)/v_0) - c_0\right)\cdot\left[1 - F\left(\psi^{-1}((c_0+\rho^*)/v_0)\right)\right].\end{aligned}$$

This completes the proof of Lemma D.1. $\square$

D.1.2. Analyzing the Class of Modified Score-Based Policies. Consider the modified score-based policy with a fixed parameter θ. Fix an arbitrary n. Assume that all agents $m \neq n$ follow the effort policy $\lambda(\cdot)$ defined in (D.4). In this proof, we refer to agent n as "focal" and other agents as "non-focal". Among the non-focal $N-1$ agents, let χ denote the number of agents whose value-to-cost ratios exceed $1/\theta$, i.e.,

$$\chi = \sum_{m \neq n} \mathbb{1}(R_m \geq 1/\theta).$$

Let $\mu(N) := (N-1)\overline{F}(1/\theta)$ and $\delta(N) := (N-1)^{2/3}$. Also, define

$$p_\mu := \mathbb{P}\left\{\chi \in (\mu(N) - \delta(N), \mu(N) + \delta(N))\right\}. \tag{D.11}$$

Throughout this proof, our focus is on $N > N' := \max\left\{1 + 1/(\overline{F}(1/\theta))^3,\ 1 + (F(1/\theta))^{3/2}\right\}$. This ensures that $\mu(N) - \delta(N) > 0$, and,

$$\frac{B}{1 + \mu(N) - \delta(N)} = \frac{\widehat{B}}{\overline{F}(1/\theta) + (1/N)F(1/\theta) - (1/N)(N-1)^{2/3}} > \frac{\widehat{B}}{\overline{F}(1/\theta)}.$$

Also, note that

$$\frac{B}{1 + \mu(N) + \delta(N)} = \frac{\widehat{B}}{\overline{F}(1/\theta) + (1/N)F(1/\theta) + (1/N)(N-1)^{2/3}} < \frac{\widehat{B}}{\overline{F}(1/\theta)}.$$

Consider the case $\widehat{B} \leq \overline{F}\left(\psi^{-1}(c_0/v_0)\right)$. Then, using θ from (D.3), we have $\frac{\widehat{B}}{\overline{F}(1/\theta)} = 1$. Thus, for $N > N'$, $B \in (1 + \mu(N) - \delta(N), 1 + \mu(N) + \delta(N))$. Consider the case $\widehat{B} > \overline{F}\left(\psi^{-1}(c_0/v_0)\right)$. Then, using θ from (D.3), we have $\frac{\widehat{B}}{\overline{F}(1/\theta)} > 1$. Thus, for $N > N'$, $B > 1 + \mu(N) - \delta(N)$.

The above two cases illustrate that it suffices to consider two intervals of B, namely $B \in (1 + \mu(N) - \delta(N), 1 + \mu(N) + \delta(N))$, and $B > 1 + \mu(N) + \delta(N)$. Thus, in the next part of this proof, we obtain lower and upper bounds on the focal agent's expected per-period utility in these two intervals of B.

Analysis for the Case: $B \in (1 + \mu(N) - \delta(N), 1 + \mu(N) + \delta(N))$. We establish a series of intermediate results, namely Lemmas D.2–D.5.

The proof of the result below extends ideas developed in proving Lemma C.1.

LEMMA D.2. *Consider any agent (say, agent l), and assume that he follows the effort policy given in* (D.4). *Then,*

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left\{S_{l,T+1} \mid R_n,\ \chi \in (\mu(N) - \delta(N), \mu(N) + \delta(N))\right\} \leq 1 - \frac{1 + \mu(N) - \delta(N)}{1 + \mu(N) + \delta(N)},$$

where R_n is the value-to-cost ratio of the focal agent n.

Proof of Lemma D.2. For the case $R_l < 1/\theta$, since agent l exerts zero effort in each period, his score remains at zero always. That is,

$$\lim_{T\to\infty} \frac{1}{T}\mathbb{E}\left\{S_{l,T+1} \mid R_n, R_l < 1/\theta,\ \chi \in (\mu(N) - \delta(N), \mu(N) + \delta(N))\right\} = 0.$$

Consider now $R_l \geq 1/\theta$. Define a sequence $\{Z_t, t \in \mathbb{N}\}$ of independent and identically distributed random variables, where, for any t,

$$Z_t = \begin{cases} 1 & \text{with probability } \frac{1+\mu(N)-\delta(N)}{1+\mu(N)+\delta(N)} \\ 0 & \text{otherwise.} \end{cases} \tag{D.12}$$

Define a sequence $\{X_{l,t}, t\in\mathbb{N}\}$ as follows:

$$X_{l,t+1} = X_{l,t} - \min(X_{l,t}, Z_t) + \theta W_{l,t}\lambda_{l,t}, \tag{D.13}$$

where $X_{l,1}=0$. Using the identity $a+b=\min(a,b)+\max(a,b)$ in (D.13), we have $X_{l,t+1}=(X_{l,t}-Z_t)^+ + \theta W_{l,t}\lambda_{l,t}$. Since $W_{l,t}\in\mathbb{R}_+$ and $\lambda_{l,t}\in\mathbb{R}_+$, we have $X_{l,t}\geq 0$ for all $t\in\mathbb{N}$. Define another sequence $\{Y_{l,t}, t\in\mathbb{N}\}$ where $Y_{l,t}=(X_{l,t}-Z_t)^+$. Then, we have:

$$Y_{l,t+1} = (X_{l,t+1} - Z_{t+1})^+ = (Y_{l,t} + \theta W_{l,t}\lambda_{l,t} - Z_{t+1})^+.$$

Using $\theta\lambda_{l,t}=1$ and $\mathbb{E}[W_{l,t}]=1$ for all t, and Theorem 2 of Loynes (1962), we have:

$$\lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left\{Y_{l,T} \mid R_n, R_l\geq 1/\theta,\ \chi\in(\mu(N)-\delta(N),\mu(N)+\delta(N))\right\} = 1-\frac{1+\mu(N)-\delta(N)}{1+\mu(N)+\delta(N)}. \tag{D.14}$$

Thus, we have:

$$\begin{aligned}
&\lim_{T\to\infty}\frac{1}{T}\mathbb{E}\left\{X_{l,T+1} \mid R_n, R_l\geq 1/\theta,\ \chi\in(\mu(N)-\delta(N),\mu(N)+\delta(N))\right\}\\
&=\lim_{T\to\infty}\left\{\mathbb{E}\left\{\frac{(X_{l,T}-Z_T)^+ + \theta W_{l,T}\lambda_{l,T}}{T}\,\middle|\, R_n, R_l\geq 1/\theta,\ \chi\in(\mu(N)-\delta(N),\mu(N)+\delta(N))\right\}\right\}\\
&=\lim_{T\to\infty}\left\{\mathbb{E}\left\{\frac{Y_{l,T} + \theta W_{l,T}\lambda_{l,T}}{T}\,\middle|\, R_n, R_l\geq 1/\theta,\ \chi\in(\mu(N)-\delta(N),\mu(N)+\delta(N))\right\}\right\}\\
&=1-\frac{1+\mu(N)-\delta(N)}{1+\mu(N)+\delta(N)}.
\end{aligned} \tag{D.15}$$

We show

$$S_{l,t}\leq_{st} X_{l,t}+1 \quad \forall t\in\mathbb{N}. \tag{D.16}$$

We use induction to prove (D.16). Since $S_{l,1}=X_{l,1}=0$, (D.16) holds for $t=1$. Suppose, for the induction hypothesis, that (D.16) holds for some $t\geq 1$. We consider the following cases:
(i) $\underline{S_{l,t}<1}$. Then, under the modified score-based policy, we have $\pi^{SB}_{l,t}=0$. Also,

$$X_{l,t+1} = X_{l,t} - \min(X_{l,t}, Z_t) + \theta W_{l,t}\lambda_{l,t} \geq \theta W_{l,t}\lambda_{l,t}.$$

Further, we have

$$S_{l,t+1} = S_{l,t} - \pi^{SB}_{l,t} + \theta W_{l,t}\lambda_{l,t} < 1+\theta W_{l,t}\lambda_{l,t} \leq 1+X_{l,t+1}.$$

(ii) $\underline{S_{l,t}\geq 1}$. Among the non-focal agents, all agents whose value-to-cost ratios are strictly below $1/\theta$ exert zero effort (see (D.4)), and their scores are always equal to zero. Given $\chi<\mu(N)+\delta(N)$, there can be at most $1+\chi<1+\mu(N)+\delta(N)$ agents (including agent l) whose scores in period t are at least 1. Further, it is given that $B>1+\mu(N)-\delta(N)$. Thus, $\pi^{SB}_{l,t}\geq_{st} Z_t$. We have:

$$\begin{aligned}
S_{l,t+1} &= S_{l,t} - \pi^{SB}_{l,t} + \theta W_{l,t}\lambda_{l,t}\\
&\leq_{st} 1+X_{l,t} - \pi^{SB}_{l,t} + \theta W_{l,t}\lambda_{l,t} \quad \text{(using induction hypothesis)}\\
&\leq_{st} 1+X_{l,t} - Z_t + \theta W_{l,t}\lambda_{l,t} \quad \text{(using } \pi^{SB}_{l,t}\geq_{st} Z_t\text{)}\\
&\leq_{st} 1+X_{l,t} - \min(X_{l,t}, Z_t) + \theta W_{l,t}\lambda_{l,t}\\
&= 1+X_{l,t+1}.
\end{aligned}$$

Thus, from the above two cases, we conclude the induction argument and hence the proof of (D.16).

Using (D.16), we have:

$$\begin{aligned}
&\lim_{T\to\infty}\frac{1}{T}\mathbb{E}\{S_{l,T+1}\mid R_n,R_l\ge 1/\theta,\ \chi\in(\mu(N)-\delta(N),\mu(N)+\delta(N))\}\\
&\le \lim_{T\to\infty}\left\{\frac{1}{T}+\frac{1}{T}\mathbb{E}\{X_{l,T+1}\mid R_n,R_l\ge 1/\theta,\ \chi\in(\mu(N)-\delta(N),\mu(N)+\delta(N))\}\right\}\\
&=1-\frac{1+\mu(N)-\delta(N)}{1+\mu(N)+\delta(N)} \qquad \text{(using (D.15))},
\end{aligned}$$

thus achieving the desired result and completing the proof of Lemma D.2. □

Lemma D.3. *The expected long-run per-period utility of the focal agent (i.e., agent n) using an arbitrary effort policy λ is bounded from above as follows: (i) If $r_n\ge 1/\theta$, then the upper bound is (v_n-c_n/θ). (ii) If $r_n<1/\theta$, then the upper bound is 0.*

Proof of Lemma D.3. This result can be proven by using the same steps as shown in the proof of Lemma B.1. Therefore, we omit its proof. □

Lemma D.4. *Recall p_μ from* (D.11). *We have $\lim_{N\to\infty}p_\mu=1$.*

Proof of Lemma D.4. Define $\overline{\chi}=\frac{1}{N-1}\chi$. Let $G_N(\cdot)$ denote the cumulative distribution function of the standardized random variable $\frac{\overline{\chi}-\overline{F}(1/\theta)}{\sqrt{F(1/\theta)\overline{F}(1/\theta)/(N-1)}}$. Let $\Phi(\cdot)$ denote the cumulative distribution function of standard normal. Then, we have:

$$\begin{aligned}
\lim_{N\to\infty}p_\mu &= \lim_{N\to\infty}\mathbb{P}\{\mu(N)-\delta(N)<\chi<\mu(N)+\delta(N)\}\\
&=\lim_{N\to\infty}\mathbb{P}\left\{-\frac{(N-1)^{2/3}}{N-1}<\overline{\chi}-\overline{F}(1/\theta)<\frac{(N-1)^{2/3}}{N-1}\right\}\\
&=\lim_{N\to\infty}\mathbb{P}\left\{-\frac{(N-1)^{1/6}}{\sqrt{F(1/\theta)\overline{F}(1/\theta)}}<\frac{\overline{\chi}-\overline{F}(1/\theta)}{\sqrt{F(1/\theta)\overline{F}(1/\theta)/(N-1)}}<\frac{(N-1)^{1/6}}{\sqrt{F(1/\theta)\overline{F}(1/\theta)}}\right\}\\
&=\lim_{N\to\infty}G_N\left(\frac{(N-1)^{1/6}}{\sqrt{F(1/\theta)\overline{F}(1/\theta)}}\right)-G_N\left(-\frac{(N-1)^{1/6}}{\sqrt{F(1/\theta)\overline{F}(1/\theta)}}\right)\\
&=\lim_{N\to\infty}\Phi\left(\frac{(N-1)^{1/6}}{\sqrt{F(1/\theta)\overline{F}(1/\theta)}}\right)-\Phi\left(-\frac{(N-1)^{1/6}}{\sqrt{F(1/\theta)\overline{F}(1/\theta)}}\right) \qquad \text{(using Central Limit Theorem)}\\
&=1.
\end{aligned}$$

This completes the proof of Lemma D.4. □

Lemma D.5. *The expected long-run per-period utility of the focal agent (i.e., agent n) using the effort policy $\lambda(\cdot)$ specified in* (D.4) *is bounded from below as follows: (i) If $r_n\ge 1/\theta$, then the lower bound is*

$$v_n\left(\frac{1+\mu(N)-\delta(N)}{1+\mu(N)+\delta(N)}\right)\cdot p_\mu-\frac{c_n}{\theta},$$

which takes the value $v_n-\frac{c_n}{\theta}$ in the limit $N\to\infty$. (ii) If $r_n<1/\theta$, then the lower bound is 0.

Proof of Lemma D.5. If $r_n < 1/\theta$, then under the effort policy in (D.4), the focal agent exerts zero effort, and is therefore, never rewarded. Thus, his expected long-run per-period utility is zero.

Consider the case $r_n \geq 1/\theta$. We have:

$$
\begin{aligned}
&\mathbb{E}\left\{\overline{\pi}_n^{SB} \,\middle|\, R_n \geq 1/\theta\right\} \\
&\geq \mathbb{E}\left\{\overline{\pi}_n^{SB} \,\middle|\, R_n \geq 1/\theta, \chi \in (\mu(N)-\delta(N), \mu(N)+\delta(N))\right\} \cdot p_\mu \quad (\text{since } \overline{\pi}_n^{SB}(\cdot) \geq 0) \\
&= \mathbb{E}\left\{-\lim_{T\to\infty} \frac{S_{n,T+1}}{T} + \theta\overline{\lambda}_n \,\middle|\, R_n \geq 1/\theta, \chi \in (\mu(N)-\delta(N), \mu(N)+\delta(N))\right\} \cdot p_\mu \quad (\text{using } S_{n,t+1} = S_{n,t} - \pi_{n,t}^{SB} + \theta \cdot q_{n,t}) \\
&\geq \frac{1+\mu(N)-\delta(N)}{1+\mu(N)+\delta(N)} \cdot p_\mu \quad (\text{using Lemma D.2}).
\end{aligned}
$$

Thus,

$$
\mathbb{E}\left[v_n \cdot \overline{\pi}_n^{SB} - c_n \cdot \overline{\lambda}_n \,\middle|\, R_n \geq 1/\theta\right] \geq v_n \left(\frac{1+\mu(N)-\delta(N)}{1+\mu(N)+\delta(N)}\right) \cdot p_\mu - \frac{c_n}{\theta}.
$$

Further, using Lemma D.4, we have $\lim_{N\to\infty} p_\mu = 1$, and thus,

$$
\lim_{N\to\infty} v_n \left(\frac{1+\mu(N)-\delta(N)}{1+\mu(N)+\delta(N)}\right) \cdot p_\mu - \frac{c_n}{\theta} = v_n - \frac{c_n}{\theta}.
$$

This completes the proof of Lemma D.5. □

Analysis for the Case: $B > 1 + \mu(N) + \delta(N)$.

Lemma D.6. *Consider any agent (say, agent l), and assume that he follows the effort policy given in* (D.4)*. Then,*

$$
\lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left\{S_{l,T+1} \mid R_n,\ \chi \in (\mu(N)-\delta(N), \mu(N)+\delta(N))\right\} = 0,
$$

where R_n is the value-to-cost ratio of the focal agent n.

Proof of Lemma D.6. The proof of this result follows similar ideas as those used in proving Lemma D.2. Therefore, for brevity, we omit a detailed proof and highlight the key changes below.

In (D.12), we replace the definition of Z_t by $Z_t = 1$.

In (D.14) and (D.15), respectively, we replace the corresponding expressions by

$$
\lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left\{Y_{l,T} \mid R_n, R_l \geq 1/\theta,\ \chi \in (\mu(N)-\delta(N), \mu(N)+\delta(N))\right\} = 0.
$$

and

$$
\lim_{T\to\infty} \frac{1}{T} \mathbb{E}\left\{X_{l,T+1} \mid R_n, R_l \geq 1/\theta,\ \chi \in (\mu(N)-\delta(N), \mu(N)+\delta(N))\right\} = 0.
$$

□

Lemma D.7. *The expected long-run per-period utility of the focal agent (i.e., agent n) using an arbitrary effort policy λ is bounded from above as follows: (i) If $r_n \geq 1/\theta$, then the upper bound is $(v_n - c_n/\theta)$. (ii) If $r_n < 1/\theta$, then the upper bound is 0.*

Proof of Lemma D.7. This result can be proven by using the same steps as shown in the proof of Lemma B.1. Therefore, we omit its proof. □

LEMMA D.8. *The expected long-run per-period utility of the focal agent (i.e., agent n) using the effort policy $\lambda(\cdot)$ specified in* (D.4) *is bounded from below as follows: (i) If $r_n \geq 1/\theta$, then the lower bound is $v_n p_\mu - c_n/\theta$, which takes the value $v_n - c_n/\theta$ in the limit $N \to \infty$. (ii) If $r_n < 1/\theta$, then the lower bound is 0.*

Proof of Lemma D.8. If $r_n < 1/\theta$, then the focal agent exerts zero effort (see (D.4)), and is therefore, never rewarded. Thus, his expected long-run per-period utility is zero.

Consider the case $r_n \geq 1/\theta$. We have:

$$
\begin{aligned}
&\mathbb{E}\left\{\overline{\pi}_n^{SB} \,\middle|\, R_n \geq 1/\theta\right\} \\
&\geq \mathbb{E}\left\{\overline{\pi}_n^{SB} \,\middle|\, R_n \geq 1/\theta, \chi \in (\mu(N)-\delta(N), \mu(N)+\delta(N))\right\} \cdot p_\mu \quad (\text{since } \overline{\pi}_n^{SB}(\cdot) \geq 0) \\
&= \mathbb{E}\left\{-\lim_{T\to\infty} \frac{S_{n,T+1}}{T} + \theta\overline{\lambda}_n \,\middle|\, R_n \geq 1/\theta, \chi \in (\mu(N)-\delta(N), \mu(N)+\delta(N))\right\} \cdot p_\mu \quad (\text{using } S_{n,t+1} = S_{n,t} - \pi_{n,t}^{SB} + \theta \cdot q_{n,t}) \\
&= p_\mu \quad (\text{using Lemma D.6}).
\end{aligned}
$$

Thus,

$$
\mathbb{E}\left[v_n \cdot \overline{\pi}_n^{SB} - c_n \cdot \overline{\lambda}_n \,\middle|\, R_n \geq 1/\theta\right] \geq v_n p_\mu - \frac{c_n}{\theta}.
$$

Further, using Lemma D.4, we have $\lim_{N\to\infty} p_\mu = 1$. This completes the proof of Lemma D.8. □

D.1.3. Concluding Steps. We now use our intermediate results from Appendix D.1.1 and Appendix D.1.2 to complete the proof of Theorem D.1. Consider the following two cases:

<u>Consider the case $\widehat{B} \leq \overline{F}(\psi^{-1}(c_0/v_0))$</u>. Then, using θ from (D.3), we have $\frac{\widehat{B}}{\overline{F}(1/\theta)} = 1$. Thus, for $N > N'$, $B \in (1+\mu(N)-\delta(N), 1+\mu(N)+\delta(N))$. Thus, using Lemmas D.3 and D.5, for any $\epsilon > 0$, there exists $N(\epsilon) > N'$ such that for all $N \geq N(\epsilon)$, the expected loss in the focal agent n's expected per-period utility from using the effort policy $\lambda(\cdot)$ in (D.4) relative to an optimal effort policy is at most ϵ.

<u>Consider the case $\widehat{B} > \overline{F}(\psi^{-1}(c_0/v_0))$</u>. Then, using θ from (D.3), we have $\frac{\widehat{B}}{\overline{F}(1/\theta)} > 1$. Thus, for $N > N'$, $B > 1+\mu(N)-\delta(N)$. Thus, using Lemmas D.3, D.5, D.7, and D.8, for any $\epsilon > 0$, there exists $N(\epsilon) > N'$ such that for all $N \geq N(\epsilon)$, the expected loss in the focal agent n's expected per-period utility from using the effort policy $\lambda(\cdot)$ in (D.4) relative to an optimal effort policy is at most ϵ.

The above two cases establish that the effort policy $\lambda(\cdot)$ specified in (D.4) constitutes an ϵ-Bayesian Nash equilibrium, completing the proof of the first part of Theorem D.1. We next evaluate the principal' expected per-period utility under the given modified score-based policy, assuming that the agents' equilibrium response is the effort policy given by (D.4).

The principal's expected per-period utility under the modified score-based policy is given as:

$$
\begin{aligned}
&\sum_{n=1}^{N} \mathbb{E}\left\{v_0\overline{\lambda}_n - c_0\overline{\pi}_n^{SB}\right\} \\
&= \sum_{n=1}^{N} \mathbb{E}\left\{v_0\overline{\lambda}_n - c_0\overline{\pi}_n^{SB} \,\middle|\, R_n \geq 1/\theta\right\} \cdot \overline{F}(1/\theta) \quad (\text{since from (D.4), } \overline{\lambda}_n = \overline{\pi}_n^{SB} = 0 \text{ if } R_n < 1/\theta) \\
&\geq \sum_{n=1}^{N} \mathbb{E}\left\{v_0\overline{\lambda}_n - c_0\theta\overline{\lambda}_n \,\middle|\, R_n \geq 1/\theta\right\} \cdot \overline{F}(1/\theta) \quad (\text{using } S_{n,t+1} = S_{n,t} - \pi_{n,t}^{SB} + \theta \cdot q_{n,t}, \text{ and } S_{n,t} \geq 0) \\
&= N(v_0/\theta - c_0)\overline{F}(1/\theta).
\end{aligned}
$$

Again, we consider the following two cases:

Consider the case $\widehat{B} \le \overline{F}\,(\psi^{-1}(c_0/v_0))$. As discussed above, we have $\frac{\widehat{B}}{\overline{F}(1/\theta)} = 1$, and thus, for $N > N'$, $B \in (1+\mu(N)-\delta(N), 1+\mu(N)+\delta(N))$. We obtain:

$$\begin{aligned}\sum_{n=1}^{N} \mathbb{E}\left\{v_0\overline{\lambda}_n - c_0\overline{\pi}_n^{SB}\right\} &\ge N(v_0/\theta - c_0)\overline{F}(1/\theta)\\ &= N\left\{v_0\psi^{-1}\left(\frac{c_0+\rho^*}{v_0}\right) - c_0\right\}\overline{F}\left(\psi^{-1}\left(\frac{c_0+\rho^*}{v_0}\right)\right) \quad \text{(using (D.3) and (D.5))}\\ &= \overline{OPT} \ge OPT \quad \text{(Lemma D.1).}\end{aligned}$$

Consider the case $\widehat{B} > \overline{F}\,(\psi^{-1}(c_0/v_0))$. Then, using θ from (D.3), we have $\frac{\widehat{B}}{\overline{F}(1/\theta)} > 1$, and thus, for $N > N'$, $B > 1+\mu(N)-\delta(N)$. We obtain:

$$\begin{aligned}\sum_{n=1}^{N} \mathbb{E}\left\{v_0\overline{\lambda}_n - c_0\overline{\pi}_n^{SB}\right\} &\ge N(v_0/\theta - c_0)\overline{F}(1/\theta)\\ &= N\left\{v_0\psi^{-1}\left(\frac{c_0}{v_0}\right) - c_0\right\}\overline{F}\left(\psi^{-1}\left(\frac{c_0}{v_0}\right)\right) \quad \text{(using (D.3))}\\ &= \overline{OPT} \ge OPT \quad \text{(using (D.5) and Lemma D.1).}\end{aligned}$$

This completes the proof of Theorem D.1. ■

Appendix E: Multi-Agent, Budget-Constrained and Private-Value Setting: Supplementary Analysis

Recall from Appendix D.1.1 that $\overline{\mathcal{P}}$ is a relaxation of the principal's problem $\mathcal{P}$ in the multi-agent, budget-constrained and private-value setting. In this appendix, we formulate problem $\overline{\mathcal{P}}$. Recall that $\overline{\mathcal{P}}$ relaxes $\mathcal{P}$ in the following three ways: (i) In $\mathcal{P}$, the reward policies in the set Π *do not explicitly depend* on the agents' attributes. Let $\Pi_0 \supseteq \Pi$ denote the set of *all* possible reward policies, including but not limited to the ones in Π. As the first step of our relaxation, we optimize over all reward policies in Π_0. (ii) Second, we relax the moral-hazard friction faced by the principal by assuming that the agent's efforts are observable to the principal. (iii) Third, we relax the budget constraint specified in (D.2) to hold in expectation (with respect to the uncertainties underlying the agents' outputs, and their privately-known value-to-cost ratios) and over the long-run. The optimal value of problem $\overline{\mathcal{P}}$ is denoted by $\overline{OPT}$.

To address the first step of our relaxation, we invoke the Revelation Principle (Myerson 1982), and, without loss of generality, restrict attention to the class of direct revelation mechanisms, which we denote by the set $\mathcal{M}$. A direct revelation mechanism $(\pi, \boldsymbol{\lambda}) \in \mathcal{M}$ is defined by a set of policies – namely, a reward policy $\pi(\cdot) \in \Pi_0$ and a collection $\boldsymbol{\lambda}$ of effort policies $\lambda_n(\cdot) \in \Lambda$, $n = 1, 2, \ldots, N$. Given that each agent n reports his attribute as (c_n, v_n), $n = 1, 2, \ldots, N$, the principal offers rewards to the agents according to the policy $\pi(\mathbf{c}, \mathbf{v})$, where $\mathbf{c} = (c_1, \ldots, c_N)$ and $\mathbf{v} = (v_1, \ldots, v_N)$, and recommends effort to each agent n according to the policy $\lambda_n(c_n, v_n)$.[5] Under any direct revelation mechanism, it is a Bayesian Nash equilibrium for all agents to report their attributes truthfully to the principal, and exert effort according to the principal's recommendation.

[5] We note that the reward policy $\pi(\cdot)$ is allowed to depend on the agents' attributes, and thus might not belong to the set Π we defined earlier.

Under the mechanism $(\pi, \boldsymbol{\lambda}) \in \mathcal{M}$, define

$$U_n\left(\hat{c}_n, \hat{v}_n, \widehat{\lambda}_n, \boldsymbol{\lambda}_{-n} \,\middle|\, (\pi, \boldsymbol{\lambda}), c_n, v_n\right) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[v_n \cdot \pi_{n,t}(\hat{c}_n, \hat{v}_n, \mathbf{C}_{-n}, \mathbf{V}_{-n}) - c_n \cdot \widehat{\lambda}_{n,t}\right] \tag{E.1}$$

as the expected long-run per-period utility of agent n who reports his attribute (c_n, v_n) as $(\hat{c}_n, \hat{v}_n)$ and exerts effort according to an effort policy $\widehat{\lambda}_n \in \Lambda$, assuming that each agent $m \neq n$ reports his attribute truthfully to the principal and follows the principal's recommended effort policy $\lambda_m := \lambda_m(c_m, v_m)$, where $\boldsymbol{\lambda}_{-n}$, $\mathbf{C}_{-n}$ and $\mathbf{V}_{-n}$, respectively, denote the vectors of the effort policies, costs and values of all agents except agent n. For notational simplicity, we have suppressed the dependence of $\pi_{n,t}$ on the history of past outputs of the agents and of $\widehat{\lambda}_{n,t}$ on the history of past outputs of the agents and agent n's attributes. Also, in the above expression for $U_n(\cdot)$, the expectation is taken with respect to the outputs of all agents, and the costs and values of all agents except agent n.

Let

$$U_n\left(\hat{c}_n, \hat{v}_n, \lambda_n(\hat{c}_n, \hat{v}_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right)$$

is the corresponding quantity for that agent when he reports his attribute as $(\hat{c}_n, \hat{v}_n)$ and obediently exerts effort according to the policy $\lambda_n(\hat{c}_n, \hat{v}_n)$ recommended by the principal, and

$$U_n\left(c_n, v_n, \lambda_n(c_n, v_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right)$$

is the corresponding quantity for that agent when he is both truthful and obedient.

Under the mechanism $(\pi, \boldsymbol{\lambda}) \in \mathcal{M}$, to ensure that the agent is both truthful and obedient, the following incentive compatibility constraints are needed:

$$U_n\left(c_n, v_n, \lambda_n(c_n, v_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right) \geq U_n\left(\hat{c}_n, \hat{v}_n, \widehat{\lambda}_n, \boldsymbol{\lambda}_{-n} \,\middle|\, (\pi, \boldsymbol{\lambda}), c_n, v_n\right) \quad \forall \hat{c}_n, \hat{v}_n, c_n, v_n, \widehat{\lambda}_n \in \Lambda, n. \tag{IC}$$

Further, to ensure that the agent receives a non-negative expected utility per-period in the long-run, the following individual rationality constraints are needed:

$$U_n\left(c_n, v_n, \lambda_n(c_n, v_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right) \geq 0 \quad \forall c_n, v_n, n. \tag{IR}$$

To address the second step of our relaxation (in which the principal faces only adverse-selection but no moral-hazard), we relax the (IC) constraints as follows:

$$U_n\left(c_n, v_n, \lambda_n(c_n, v_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right) \geq U_n\left(\hat{c}_n, \hat{v}_n, \lambda_n(\hat{c}_n, \hat{v}_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right) \quad \forall \hat{c}_n, \hat{v}_n, c_n, v_n, n. \tag{$\overline{IC}$}$$

To address the third step of our relaxation, we relax the budget constraint (D.2) as follows:

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} \mathbb{E}[\pi_{n,t}(\mathbf{C}, \mathbf{V})] \leq B. \tag{E.2}$$

The relaxed problem $\overline{\mathcal{P}}$ is now formulated as follows:

$$\begin{aligned} \overline{OPT} \;:=\; \max_{(\pi,\boldsymbol{\lambda})\in\mathcal{M}} \quad & \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} \mathbb{E}\Big[v_0 \cdot W_{n,t} \cdot \lambda_{n,t}(C_n, V_n) \;-\; c_0 \cdot \pi_{n,t}(\mathbf{C}, \mathbf{V})\Big] \\ \text{s.t.} \quad & U_n\left(c_n, v_n, \lambda_n(c_n, v_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right) \end{aligned}$$

$$\geq U_n\left(\hat{c}_n, \hat{v}_n, \lambda_n(\hat{c}_n, \hat{v}_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right) \quad \forall \hat{c}_n, \hat{v}_n, c_n, v_n, n, \tag{$\overline{IC}$}$$

$$U_n\left(c_n, v_n, \lambda_n(c_n, v_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right) \geq 0 \quad \forall c_n, v_n, n, \tag{IR}$$

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} \mathbb{E}[\pi_{n,t}(\mathbf{C}, \mathbf{V})] \leq B. \tag{E.2}$$

We now use standard techniques in mechanism design to develop a *reduced* form of problem $\overline{\mathcal{P}}$. Consider an arbitrary direct revelation mechanism $(\pi, \boldsymbol{\lambda}) \in \mathcal{M}$. Using (E.1), the expected long-run per-period utility of agent n who reports his attribute (c_n, v_n) as $(\hat{c}_n, \hat{v}_n)$, and exerts effort according to the policy $\lambda_n(\hat{c}_n, \hat{v}_n)$ recommended by the principal is:

$$\begin{aligned} U_n\left(\hat{c}_n, \hat{v}_n, \lambda_n(\hat{c}_n, \hat{v}_n), \boldsymbol{\lambda}_{-n} \mid (\pi, \boldsymbol{\lambda}), c_n, v_n\right) &= \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[v_n \cdot \pi_{n,t}(\hat{c}_n, \hat{v}_n, \mathbf{C}_{-n}, \mathbf{V}_{-n}) - c_n \cdot \lambda_{n,t}(\hat{c}_n, \hat{v}_n)\right] \\ &= v_n \cdot \overline{\pi}_n(\hat{c}_n, \hat{v}_n) - c_n \cdot \overline{\lambda}_n(\hat{c}_n, \hat{v}_n), \end{aligned} \tag{E.3}$$

where

$$\overline{\pi}_n(\hat{c}_n, \hat{v}_n) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}\left[\pi_{n,t}(\hat{c}_n, \hat{v}_n, \mathbf{C}_{-n}, \mathbf{V}_{-n})\right] \quad \text{and} \quad \overline{\lambda}_n(\hat{c}_n, \hat{v}_n) := \lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \mathbb{E}[\lambda_{n,t}(\hat{c}_n, \hat{v}_n)].$$

Using (E.3), the constraint ($\overline{IC}$) reduces to:

$$v_n \cdot \overline{\pi}_n(c_n, v_n) - c_n \cdot \overline{\lambda}_n(c_n, v_n) \geq v_n \cdot \overline{\pi}_n(\hat{c}_n, \hat{v}_n) - c_n \cdot \overline{\lambda}_n(\hat{c}_n, \hat{v}_n) \quad \forall \hat{c}_n, \hat{v}_n, c_n, v_n, n.$$

Similarly, we rewrite the (IR) constraint as:

$$v_n \cdot \overline{\pi}_n(c_n, v_n) - c_n \cdot \overline{\lambda}_n(c_n, v_n) \geq 0 \quad \forall c_n, v_n, n,$$

and the objective function in $\overline{\mathcal{P}}$ as:

$$\begin{aligned} &\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} \mathbb{E}\Big[v_0 \cdot W_{n,t} \cdot \lambda_{n,t}(C_n, V_n) \; - \; c_0 \cdot \pi_{n,t}(\mathbf{C}, \mathbf{V})\Big] \\ =&\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} \sum_{n=1}^{N} \{v_0 \cdot \mathbb{E}[W_{n,t}] \cdot \mathbb{E}[\lambda_{n,t}(C_n, V_n)] \; - \; c_0 \cdot \mathbb{E}[\pi_{n,t}(\mathbf{C}, \mathbf{V})]\} \\ =&\sum_{n=1}^{N} \mathbb{E}\left[v_0 \cdot \overline{\lambda}_n(C_n, V_n) - c_0 \cdot \overline{\pi}_n(C_n, V_n)\right], \end{aligned}$$

where, in the first equality, we use the independence of $W_{n,t}$ across t and also the independence between $W_{n,t}$ and (C_n, V_n), and in the second equality, we use $\mathbb{E}[W_{n,t}] = 1$ for all (n, t).

Using the steps described above, the relaxed problem $\overline{\mathcal{P}}$ reduces to the following optimization problem:

$$\begin{aligned} \max_{\{\overline{\pi}_n(\cdot), \overline{\lambda}_n(\cdot)\} \ \forall n} \quad & \sum_{n=1}^{N} \mathbb{E}\left[v_0 \cdot \overline{\lambda}_n(C_n, V_n) - c_0 \cdot \overline{\pi}_n(C_n, V_n)\right] \\ \text{s.t.} \quad & v_n \cdot \overline{\pi}_n(c_n, v_n) - c_n \cdot \overline{\lambda}_n(c_n, v_n) \geq v_n \cdot \overline{\pi}_n(\hat{c}_n, \hat{v}_n) - c_n \cdot \overline{\lambda}_n(\hat{c}_n, \hat{v}_n) \quad \forall \hat{c}_n, \hat{v}_n, c_n, v_n, n, \\ & v_n \cdot \overline{\pi}_n(c_n, v_n) - c_n \cdot \overline{\lambda}_n(c_n, v_n) \geq 0 \quad \forall c_n, v_n, n, \\ & \sum_{n=1}^{N} \mathbb{E}[\overline{\pi}_n(C_n, V_n)] \leq B. \end{aligned} \tag{E.2}$$

Let $r_n := v_n/c_n$ denote the value-to-cost ratio of agent n. Further, note that the agent's attribute (c_n, v_n) affects the above optimization problem *only through the ratio* r_n. Following Theorem 3.1 in Balseiro et al. (2019a), without loss of generality, the above optimization problem can be recast as a single-dimensional mechanism-design problem as follows:

$$\max_{\{\overline{\pi}_n(\cdot),\overline{\lambda}_n(\cdot)\}\ \forall n} \quad \sum_{n=1}^{N} \mathbb{E}\left[v_0 \cdot \overline{\lambda}_n(R_n) - c_0 \cdot \overline{\pi}_n(R_n)\right] \tag{E.4}$$

$$\text{s.t.} \quad r_n \cdot \overline{\pi}_n(r_n) - \overline{\lambda}_n(r_n) \geq r_n \cdot \overline{\pi}_n(\hat{r}_n) - \overline{\lambda}_n(\hat{r}_n) \quad \forall \hat{r}_n, r_n, n, \tag{E.5}$$

$$r_n \cdot \overline{\pi}_n(r_n) - \overline{\lambda}_n(r_n) \geq 0 \quad \forall r_n, n, \tag{E.6}$$

$$\sum_{n=1}^{N} \mathbb{E}[\overline{\pi}_n(R_n)] \leq B. \tag{E.7}$$

Let $y_n(r_n) := r_n \cdot \overline{\pi}_n(r_n) - \overline{\lambda}_n(r_n)$ for any $r_n \in \mathbb{R}_+$. Following the standard procedure for solving single-dimensional mechanism design problems, we note that (E.5) and (E.6) hold if and only if the following conditions hold (see Lemma 2 of Myerson 1981):

$$y_n(r_n) = y_n(0) + \int_0^{r_n} \overline{\pi}_n(r')dr' \quad \forall r_n, n, \tag{E.8}$$

$$\overline{\pi}_n(r) \geq \overline{\pi}_n(r') \quad \forall r \geq r', n, \tag{E.9}$$

$$y_n(0) \geq 0 \quad \forall n. \tag{E.10}$$

Thus, we have

$$\begin{aligned}
\overline{OPT} &= \max_{\{\overline{\pi}_n(\cdot),\overline{\lambda}_n(\cdot)\}\ \forall n} \quad \sum_{n=1}^{N} \mathbb{E}\left[v_0 \cdot \overline{\lambda}_n(R_n) - c_0 \cdot \overline{\pi}_n(R_n)\right] \quad \text{s.t. (E.5)–(E.7)} \\
&= \max_{\{\overline{\pi}_n(\cdot),\overline{\lambda}_n(\cdot)\}\ \forall n} \quad \sum_{n=1}^{N} \mathbb{E}\left[v_0 \cdot \overline{\lambda}_n(R_n) - c_0 \cdot \overline{\pi}_n(R_n)\right] \quad \text{s.t. (E.7)–(E.10)}, \\
&= \max_{\{\overline{\pi}_n(\cdot),\overline{\lambda}_n(\cdot)\}\ \forall n} \quad \sum_{n=1}^{N} \mathbb{E}\left[v_0 \cdot \psi(R_n)\overline{\pi}_n(R_n) - v_0 y_n(0) - c_0 \cdot \overline{\pi}_n(R_n)\right] \quad \text{s.t. (E.7)–(E.10)},
\end{aligned}$$

where the last equality follows from substituting (E.8) in (E.4), using the standard exchange of integrals and the definition of $\psi(\cdot)$ in (16) (see, e.g., Myerson 1981).

Appendix F: Connection Between Limited-Term and Score-Based Reward Policies

In this appendix, we present a theoretically-grounded connection between limited-term and score-based policies. This connection also offers an understanding of why the class of score-based policies is guaranteed to always contain an optimal solution for the principal.

A Brief Summary: Score-based policies are, in fact, closely connected to limited-term (in short, LT) policies, which are popular non-monetary reward policies in practice. To establish this connection, we introduce a new, auxiliary class of policies, which we refer to as $\widehat{LT}$.[6] This new class acts as a "bridge" between score-based policies and LT policies. On the one hand, a score-based policy is structurally similar to an $\widehat{LT}$ policy,

[6] Our sole purpose in defining the class $\widehat{LT}$ is to illustrate the connection between LT and score-based policies.

with the only difference being in the amount of non-monetary reward the agent is given based on his past performance. On the other hand, an LT policy is similar to an $\widehat{LT}$ policy, with the only difference being that in the LT policy, the agent has to sometimes forego rewards that she qualifies for, whereas in the $\widehat{LT}$ policy, the agent is always able to carry forward or accumulate rewards over time.

This connection also highlights a specific property of score-based policies that is missing from LT policies, namely *the ability to carry forward any unused rewards that have been earned.* This property helps the score-based class to achieve optimality in problem $\mathcal{P}$. In other words, while the class of LT policies is not guaranteed to always contain an optimal policy, the class of score-based policies always contains one. □

We now discuss these connections in more detail. Recall the definition of limited-term policies (Section 4). In an LT policy with parameters (t_1, t_2, m), in any period t, if the agent's cumulative output over the past t_1 periods (including the current period) is greater than or equal to the threshold m, then the agent qualifies for the non-monetary reward in each of the next t_2 periods. Note, however, that the agent can receive at most one reward in any period.[7] Thus, *in any given period, if the agent qualified for the reward multiple times due to passing multiple evaluations in the past, he receives only one reward.*

To establish a connection between limited-term and score-based policies, let us define another class of policies, denoted by $\widehat{LT}$. Like LT policies, $\widehat{LT}$ policies are also indexed by the parameters (t_1, t_2, m). As with an LT policy, in an $\widehat{LT}$ policy with parameters (t_1, t_2, m), in any period t, if the agent's cumulative output over the past t_1 periods exceeds m, then the agent receives the reward in each of the next t_2 periods. However, unlike the LT policy, in any given period of the $\widehat{LT}$ policy, if the agent qualified for a reward multiple times due to passing multiple evaluations in the past, then he receives one reward in that period and can *carry forward* any remaining rewards to future periods. Formally, an $\widehat{LT}$ policy is defined as follows: Let A_t (where the letter A stands for account) represent the number of rewards accumulated by the agent until (but excluding) period t from successfully passing reviews. Then, for any period $t \geq 1$, we have:

$$\pi_t^{\widehat{LT}} = \begin{cases} 1 & \text{if } A_t \geq 1, \\ 0 & \text{otherwise,} \end{cases} \tag{F.1}$$

$$\text{where} \quad A_{t+1} = A_t - \pi_t^{\widehat{LT}} + t_2 \cdot \mathbb{1}\left[\sum_{s=t-t_1+1}^{s=t} \lambda_s W_s \geq m\right],$$

where $A_1 = 0$, λ_s denotes the effort exerted by the agent, and W_s denotes the random variable influencing the agent's output in period s.

We briefly explain the definition (F.1) above in words. For any t, the agent is rewarded in period t, i.e., $\pi_t^{\widehat{LT}} = 1$, if and only if he has at least one reward in his account at the beginning of period t, i.e., $A_t \geq 1$. Thus, the agent's account A_t decreases by the amount $\pi_t^{\widehat{LT}}$ in period t. If the agent's cumulative output over the past t_1 number of periods (i.e., periods $t - t_1 + 1$ through t) exceeds m, then he passes the review in period t and his account increases by an amount t_2.

We now examine how $\widehat{LT}$ policies are connected to both LT and score-based policies.

[7] For example, a faculty can receive at most one teaching-load reduction in an academic year, or a supplier can receive at most one Supplier-of-the-Year award in an year.

The Connection Between LT and $\widehat{LT}$ Policies:

We discuss the following illustrative example to explain the difference between an LT and an $\widehat{LT}$ policy, both with parameters (t_1, t_2, m):

> **Example:** Suppose that $t_1 = 1$ and $t_2 = 2$. Consider any two consecutive periods t and $t+1$ and assume that the agent delivers a sufficiently high output (i.e., larger than m) in both periods (for simplicity, suppose that the agent's output is 0 before period t and after period $t+1$). Consider the LT policy with parameters (t_1, t_2, m). Since the output in period t exceeds m and since $t_2 = 2$, the agent qualifies for the reward in each of the next $t_2 = 2$ periods, i.e., periods $t+1$ and $t+2$. Since the output in period $t+1$ also exceeds m, the agent also qualifies for the reward in each of the next two periods, i.e., periods $t+2$ and $t+3$. Thus, the agent qualifies for the reward two times in period $t+2$ (one from the review in period t and another from the review in period $t+1$). However, since the agent can receive at most one reward in any given period, he has to *forego the additional reward he qualified for* in period $t+2$. Thus, by passing review in periods t and $t+1$, the agent receives rewards in three periods, i.e., periods $t+1$, $t+2$ and $t+3$. Consider now the $\widehat{LT}$ policy with parameters (t_1, t_2, m). As in the LT policy, in the $\widehat{LT}$ policy too, the agent qualifies for the reward two times in period $t+2$, one from the review in period t and another from the review in period $t+1$. However, unlike the LT policy, under the $\widehat{LT}$ policy, the agent receives one reward in period $t+2$ and *carries forward the additional reward* to the next period, i.e., period $t+3$. As a result, the agent qualifies for two rewards in period $t+3$, one from passing the review in period $t+1$ and another that is carried forward from period $t+2$. Again, since the agent can receive at most one reward in a period, he receives one reward in period $t+3$ and one reward in period $t+4$. Thus, by passing the reviews in periods t and $t+1$, the agent can earns rewards in four periods, i.e., periods $t+1$ through $t+4$. In this manner, while the agent may have to forego rewards under an LT policy in some cases, he is able to avoid losing rewards under the $\widehat{LT}$ policy by carrying forward rewards over time. □

Claim 1 below demonstrates an instance of problem $\mathcal{P}$ for which (i) *every* LT policy is suboptimal and (ii) a specific $\widehat{LT}$ policy is optimal. The proof of Claim 1 is provided in Appendix F.1.

CLAIM 1. *Consider the instance of problem $\mathcal{P}$ in which $W_t = 2$ with probability $1/2$ and $W_t = 0$ otherwise. Further, $OPT_B = v_0 r - c_0 > 0$, where OPT_B is obtained in* (3). *Then, we have the following two results: (a) Every LT policy is sub-optimal. (b) The $\widehat{LT}$ policy with parameters $t_1 = 1$, $t_2 = 2$ and $m = 2 \cdot r$ is optimal.*

We briefly explain the significance of Claim 1. Recall that the only difference between an $\widehat{LT}$ and an LT policy is that, in the former, the agent is allowed to carry forward rewards over time. By using an instance of problem $\mathcal{P}$, Claim 1 illustrates that this property is indeed necessary to attain optimality in problem $\mathcal{P}$. We note that this property is also present in score-based reward policies, as we discuss below.

The Connection Between $\widehat{LT}$ and Score-Based Policies:

We recall the definition of a score-based reward policy from Section 6. Specifically, the score-based policy maintains a score for the agent through time, and dynamically updates this score according to (12) based on the agent's performance. Comparing (F.1) and (12), it is clear that the only difference between an $\widehat{LT}$ policy and a score-based policy is the amount by which the account under the $\widehat{LT}$ policy and the score under the score-based policy increase with time. Specifically, under the $\widehat{LT}$ policy, the account increases by an amount $t_2 \cdot \mathbb{1}\left[\sum_{s=t-t_1+1}^{s=t} \lambda_s W_s \geq m\right]$ in period t. That is, if the agent passes the review in period t (i.e., the value of the binary variable in the preceding expression equals 1), then his account increases by an amount t_2. In contrast, under the score-based policy, the score increases by an amount $\theta\lambda_t W_t$ in period t.

Under the $\widehat{LT}$ policy, the rewards earned in the past carry forward to future periods via the account A_t defined in (F.1). In a similar fashion, the score in the score-based policy can be viewed as "credits" the agent receives as a function of his performance – here too, the credits earned in the past carry forward to the future via the score S_t defined in (12).

F.1. Proof of Claim 1

The proof of part (a) of Claim 1 is the same as the proof of Theorem 2 that is provided in Appendix B. We now prove part (b) of the claim. For the $\widehat{LT}$ policy with parameters $t_1 = 1$, $t_2 = 2$ and $m = 2 \cdot r$, using (F.1), we have:

$$\begin{aligned}\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}[\pi_t^{\widehat{LT}}] &= \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[A_t - A_{t+1} + 2\cdot \mathbb{1}\left[W_t\lambda_t \geq m\right]\right] \\ &\leq \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} 2\mathbb{P}[W_t\lambda_t \geq m] \quad \text{(since } A_1 = 0 \text{ and } A_t \geq 0\ \forall t).\end{aligned}$$

Thus, the expected long-run per-period utility of the agent is:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[v\cdot \pi_t^{\widehat{LT}} - c\cdot\lambda_t\right] \leq \max_{\{\lambda_t,\ \forall t\}} \quad \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T}\left[v\cdot 2\mathbb{P}[W_t\lambda_t \geq m] - c\lambda_t\right].$$

Since the optimization problem in the right-hand-side of the above inequality is separable in t, we can maximize the given expression for any t. For the given distribution of W_t, since $m > 0$, we have

$$\mathbb{P}[W_t\lambda_t \geq m] = (1/2)\mathbb{1}[0 \geq m] + (1/2)\mathbb{1}[2\lambda_t \geq m] = (1/2)\mathbb{1}[2\lambda_t \geq m].$$

Thus,

$$v\cdot 2\mathbb{P}[W_t\lambda_t \geq m] - c\lambda_t = \begin{cases} -c\lambda_t & \lambda_t < m/2 \\ v - c\lambda_t & \lambda_t \geq m/2. \end{cases}$$

Thus, the value of λ_t which maximizes $v\cdot 2\mathbb{P}[W_t\lambda_t \geq m] - c\lambda_t$ is given as:

$$\lambda_t = \begin{cases} m/2 & v/c \geq m/2 \\ 0 & \text{otherwise.} \end{cases} \tag{F.2}$$

Using (F.2), the upper bound on the agent's expected utility is obtained as:

$$\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[v\cdot \pi_t^{\widehat{LT}} - c\cdot\lambda_t\right] \leq \begin{cases} v - cm/2 & v/c \geq m/2 \\ 0 & \text{otherwise.} \end{cases} \tag{F.3}$$

Next, we show that the above bound is tight if the agent follows the effort policy given in (F.2). To this end, we need the following intermediate lemma.

Lemma F.1. *Under the* $\widehat{LT}$ *policy with parameters* $t_1 = 1$, $t_2 = 2$ *and* $m = 2r$, *suppose that the agent follows the effort policy specified in* (F.2). *Then, we have:*

$$\lim_{T\to\infty} \frac{\mathbb{E}[A_{T+1}]}{T} = 0.$$

The proof of Lemma F.1 mirrors that of Lemma B.2 in Appendix B, with the modifications that S_t is replaced by A_t, $\lambda_t = 1/\theta$ is replaced by $\lambda_t = m/2$, and $\theta\lambda_t W_t$ is replaced by $2 \cdot \mathbb{1}[W_t\lambda_t \geq m]$. Therefore, for brevity, we omit the proof of Lemma F.1.

Given that the agent follows the effort policy specified in (F.2), we have:

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[v \cdot \pi_t^{\widehat{LT}} - c \cdot \lambda_t\right] &= v \cdot \left(- \lim_{T\to\infty} \frac{\mathbb{E}[A_{T+1}]}{T} + \lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} 2 \cdot \mathbb{P}[W_t\lambda_t \geq m]\right) - c\overline{\lambda} \quad \text{(using (F.1))}\\
&= v \cdot \left(\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} 2 \cdot \mathbb{P}[W_t\lambda_t \geq m]\right) - c\overline{\lambda} \quad \text{(using Lemma F.1)}\\
&= \begin{cases} v - cm/2 & v/c \geq m/2 \\ 0 & \text{otherwise,} \end{cases}
\end{aligned}$$

and thus, the upper bound obtained on the agent's expected utility in (F.3) is tight. Thus, the effort policy given in (F.2) is an optimal response of the agent under the given $\widehat{LT}$ reward policy. Given $m = 2r$, the principal's expected long-run per-period utility under the given policy is:

$$\begin{aligned}
\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \mathbb{E}\left[v_0 W_t\lambda_t - c_0\pi_t^{\widehat{LT}}\right] &= v_0 \cdot \overline{\lambda} - c_0 \cdot \left(\lim_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} 2 \cdot \mathbb{P}[W_t\lambda_t \geq m]\right)\\
&= v_0 m/2 - c_0\\
&= OPT_B \quad \text{(using } m = 2r \text{ and (3)).}
\end{aligned}$$

The above equality together with the fact that OPT_B is an upper bound on the principal's optimal value OPT in problem $\mathcal{P}$, completes our proof of Claim 1. ■