Incentives and Burnout: Dynamic Compensation Design With Effort Cost Spillover

Rob Waiser Juan Dubra Jean-Pierre Benoît

Preliminary draft – Please do not circulate

1 Introduction

According to a recent survey, "95 percent of HR leaders admit that employee burnout is sabotaging workforce retention", with the top three contributing factors cited as "unfair compensation, unreasonable workload, and too much overtime / after-hours work". Furthermore, even among employees who stay with their firms, burnout or fatigue can undermine engagement, sap productivity and fuel absenteeism (Kronos, 2017).

According to the World Health Organization (2018), burnout is "a syndrome conceptualized as resulting from chronic workplace stress that has not been successfully managed. It is characterised by three dimensions: 1) feelings of energy depletion or exhaustion; 2) increased mental distance from one's job, or feelings of negativism or cynicism related to one's job; and 3) a sense of ineffectiveness and lack of accomplishment." Numerous studies (e.g., Cordes & Dougherty, 1993; Maslach et al., 2001; Schaufeli & Bakker, 2004) have found that one of the primary drivers of burnout is a heavy workload (often referred to as "role overload" or "quantitative overload"), which is typically treated as a fixed characteristic of a job. In practice, however, workers make choices about how hard they work, particularly when their effort is not directly observed. Employers, in turn, try to affect those choices using incentives, targets, consequences, etc.

Consider the following example, from Sullivan (2018):

A few years ago, at her annual review, a friend was encouraged to pursue 'stretch' goals, and told that she could do more than she ever thought possible. Inspired, she worked hard to achieve these new goals, learning and growing along the way. However, she also sacrificed a lot by working long hours, as well as over weekends.

At the next review, her boss said: 'I knew you could do it! Let's bump up your goals again.' Still in the glow of success she agreed. This time it was harder to sacrifice but she scraped by. The following year, more fatigued, she was not so welcoming of her boss's praise and pitch for even higher targets again. The fourth year, overwhelmed and dissatisfied, she quit.

This story highlights the fact that the cost of an employee's effort (e.g. fatigue, opportunity cost of time) spills over from one period to the next. Burnout can be thought of as resulting from the accumulation of this spillover. In the example, the friend's exertion in the pursuit of one year's goal affects her willingness and/or ability to work hard in the next year, as she progresses from "inspired" to "scraping by" to "fatigued" and ultimately "overwhelmed".[1]

From a firm's standpoint, this 'effort cost spillover' implies that employees may need to be compensated differently over time, to account for the accumulating effects of their effort. If an employee is incentivized to work very hard in one period, as in our example, she may require even stronger incentives in the next period in order to maintain her motivation or keep her from quitting. In that case, is it optimal for the firm to reduce the employee's incentives in the first period in order to reduce her effort cost in the next? Or is it better to offer strong incentives when the employee is 'fresh', then lower them – or perhaps let her leave the firm – when fatigue sets in?

These questions are not addressed by existing incentive models. A typical approach to modeling incentive design for workers whose actions are unobserved is to use a principal-agent model (e.g. Holmström, 1979; Basu et al., 1985) that represents a single period, with some exceptions discussed below. The implicit assumption is that the firm can follow single-period

[1]The story also illustrates 'target ratcheting', in which the firm responds to an employee's strong performance by raising goals or expectations. Ratcheting has been well studied by researchers in both economics (e.g. Weitzman, 1980; Laffont & Tirole, 1988) and accounting (e.g. Bouwens & Kroos, 2011) and is not the focus of this paper.

reasoning at each sage independently. However, this does not allow the principal or the agent to account for any spillovers in the agent's effort cost from one period to the next.

We develop a dynamic two-period principal-agent model with a risk-neutral agent (worker) who experiences effort cost spillover between periods. Specifically, the worker's effort cost in the second period is increasing in both her second-period effort *and* her first-period effort. We explore the optimal contract design over time and consider the connection between burnout and contract design. We treat 'workload' as a choice made by the worker in response to her contract (and her expectations about future contracts), rather than a fixed job characteristic.

Cordes et al. (1997) note that "Burnout is a developmental process. There is no on-off switch, no clearly defined moment at which an employee announces 'I am burned out'." Accordingly, the effort cost spillover effect in our model increases continuously with the worker's effort. With that said, our model results do suggest the possibility of a particular type of outcome that we call a "burnout equilibrium". In layman's terms, it is intuitive to think of someone as 'burned out' when they are no longer willing to do their job as a result of their past effort, as in the above example. However, this is rather imprecise, particularly when we endogenize contract design. Would the 'friend' in that example have been willing to continue working if she received a pay raise of $1 million? $10 million? $100 million?

We define a 'burnout equilibrium' as an equilibrium in which the firm incentivizes the worker to exert so much effort in the first period that she is unwilling to accept any contract that the firm finds profitable in the second period. Put differently, she works so hard initially that in order to continue working, she would need an exorbitant remuneration that the firm is unwilling to provide. In our analysis, we restrict ourselves to situations in which the efficient outcome – i.e., the outcome which maximizes the total long-term surplus – requires the worker to stay with the firm and exert positive effort in both periods. Among other things, we are interested in whether a burnout equilibrium can arise even under those conditions.

With no spillovers, and a risk-neutral worker, the firm can always capture the efficient two-period surplus by offering a franchise contract[2] in each period. This result obtains whether the firm and worker are forward-looking or myopic, whether or not the firm can commit to future contracts, and independently of modeling details such as whether the worker's effort choice is continuous or discrete. The simplicity and robustness of this finding in the no-

[2]In a franchise contract, the worker is paid the net revenue she produces and charged a "franchise fee".

spillover case is one reason we choose to model the worker as risk-neutral in our setting. As we will see, the presence of spillovers produces significant changes in the analysis and in our understanding of incentive contracts.

Suppose there are spillovers and the actors behave myopically, both making decisions in each period without looking ahead. Unsurprisingly, the firm than earns less than the efficient total profits. In the first period, the worker exerts the single-period efficient effort but this may be inefficiently high considering both periods. In general, the worker exerts too much effort in period 1 and, due to the spillover, too little effort in period 2. This may be somewhat of a benchmark finding, given that we usually expect firms to take a long-term perspective.

Suppose instead that the firm is forward-looking but the worker is myopic. This situation corresponds to an experienced firm and a naive worker who does not properly understand the long run impact of effort on her. The firm can correct for the worker's short-sightedness by offering a first-period contract with reduced incentives that lead her to optimally restrain her first-period effort. In this way, the firm can always achieve its first-best outcome.

Now suppose that the firm and worker are both forward-looking. Somewhat surprisingly, the firm may no longer be able to achieve its first-best. We might have expected that the firm could obtain its first-best outcome by simply offering two franchise contracts. After all, a forward-looking worker should internalize the spillover externality and lower her first-period effort on her own. Or we might have expected that offering a first-period contract with reduced incentives could achieve the first-best. The rub, however, is that the worker anticipates that the firm will offer a second-period contract that leaves her with zero surplus in that period. As a consequence, if the worker is offered a franchise contract in the first period, she may choose to over-exert herself in that period and quit in the second (thereby avoiding the spillover cost). If the worker is instead offered a first-period contract with reduced incentives, she may choose to shirk in that period to benefit from a lower second-period effort cost. Thus, in sharp contrast with the no-spillover case, the firm may be unable to obtain first-best profits. Moreover, we find that the firm's equilibrium strategy may be to induce the worker to burn herself out, even though she cannot be replaced in the second period and it is efficient for her to stay with the firm and continue exerting effort. This burnout equilibrium can be quite inefficient, yielding little more than half of the first-best surplus.

The firm is harmed by the fact that the worker anticipates a relatively unattractive second-period contract offer. This obstacle can be surmounted if the firm is willing, and able, to commit to an incentive plan for *both periods at once.* The firm can then always achieve its first-best outcome by offering a period 1 contract that yields the worker a negative surplus and committing to a period 2 contract that yields a positive surplus, with zero total surplus for the worker. This commitment need not be a formal, multi-period contract. It could, for instance, be a 'reputational commitment' in which the worker trusts the firm to offer a particular period two contract, with the firm reliably doing so to avoid 'punishment' by workers in the future.

1.1 Literature review

As noted above, much of the theoretical literature on incentive plan design for workers in the last four decades is based in agency theory. The underlying principal-agent models assume an outcome (e.g. output, sales, profits) that is driven by the effort of the agent, but with a stochastic component, reflecting uncertainty in the production function. While these models have been adapted to a wide range of contexts (e.g. Lal & Staelin, 1986; Joseph & Thevaranjan, 1998; Godes, 2004), they typically consider only a single time period, wherein the firm (principal) offers a contract, then the worker (agent) chooses whether to accept and, if so, chooses her effort level for the period to which the contract applies.

There is, however, a subset of the compensation literature that considers multiple periods, including a mix of empirical and theoretical research. On the empirical side, Banker et al. (2001) provide a study of performance improvements following the implementation of a "pay-for-performance" compensation plan, using data from multiple periods to explore the relative strength of the "selection effect" and the "effort effect" of incentives. Steenburgh (2008) provides an empirical study of the effects of lump-sum bonuses on salesperson behaviour, including the potential for manipulation of order timing. A number of researchers have used multi-period studies to explore the effects of "target ratcheting", the practice of basing targets for one period (at least in part) on performance in the previous period. See Indjejikian et al. (2014) for a review of both theory and empirical research on ratcheting.

Multi-period worker compensation models have been considered somewhat more extensively in the theoretical literature, but the firm's and worker's utilities are generally assumed to be independent or 'time-separable' across periods (e.g. Mantrala et al., 1997; Gershkov & Perry, 2012; Jerath & Long, 2020). Similarly, dynamic moral hazard models in which decisions are made in continuous time (e.g. Holmström & Milgrom, 1987; Lal & Srinivasan, 1993; Demarzo & Sannikov, 2017) typically assume that preferences remain consistent over time. In our model, the agent's utility – specifically her effort cost – depends on her action in the previous period, introducing an important new dynamic. Schöttner (2017) allows for similar "cost externalities" in a two-period principal-agent model, but the firm only observes outcomes (and determines compensation) at the end of the second period. The agent's effort options are binary and the firm always prefers the higher option in each period. We relax those assumptions, allowing the firm to compensate the agent after each period and considering a range of optimal effort choices. Furthermore, while Schöttner (2017) focuses on whether the firm should offer commission- or bonus-based compensation, we explore how and when the firm can achieve its best possible outcome and the conditions under which it prefers to burn agents out.

There is also a subset of the dynamic moral hazard literature that focuses on intertemporal risk sharing and how past outputs affect the curvature of the agent's utility function (e.g. Rogerson, 1985; Lambert, 1983). The effect of savings on the optimal incentive schemes have also been analyzed by Fudenberg et al. (1990) and the literature that followed (in particular, the literature that considers hidden savings, such as Ábrahám & Pavoni (2008) and the references therein). Again, in those literatures the agent's cost of effort remains consistent over time, with dynamics affecting her accumulated wealth (and thus utility). In our model, we focus on a risk-neutral agent in order to isolate the cause of inefficiency, so the insights from those subsets of the literature do not apply.

The model that is perhaps most similar to our own is that used by Dearden & Lilien (1990). That paper focuses on "production learning", through which manufacturing costs decline with cumulative volume (i.e., with production experience). When that occurs, firms have an incentive to increase sales, and therefore production, in one period in order to decrease costs in the next. The authors use a two-period model to explore the use of sales force compensation to achieve this objective, optimizing long-term discounted profits in the presence of production learning. Again, however, their model assumes that the agent's utility with respect to income and effort is identical and independent across periods. As a result, their agent does not benefit from a forward-looking strategy, although the firm does. Perhaps not surprisingly, then, their results look quite similar to our own (although inverted, because early effort lowers later costs in their model and raises them in ours) if the agent were assumed to be myopic and the firm forward-looking. However, our model introduces new findings when the agent is forward-looking, adjusting her early effort in anticipation of its effect on her future utility.

Lastly, there is a substantial literature on the subject of employee burnout. The commonly accepted definition in that research is similar to that of the WHO, focusing on three components: emotional exhaustion, depersonalization / cynicism, and diminished personal accomplishment (Cordes & Dougherty, 1993). As noted above, studies have consistently found that "work overload" is a significant driver of emotional exhaustion and consequently burnout, but workload is typically treated as a fixed characteristic of a job and the link between workload and incentives is rarely considered. One notable exception is Habel et al. (2021), an empirical study of the effects of variable compensation on salesperson health, including emotional exhaustion. Habel et al. find that variable compensation (as a share of total compensation) is positively associated with both performance and emotional exhaustion, which is consistent with our model assumptions. Ours is believed to be the first use of game theory to study burnout. While most of the existing research focuses on the process of burnout or its three components, we are interested in the relationships between burnout, incentives, and turnover. We treat workload as the result of choices made by the firm and the worker, rather than a fixed characteristic.

The remainder of this paper is structured as follows: In the next section, we introduce the model to be analyzed. Next, we derive and discuss the results of the model. Finally, we review the implications of our findings and discuss related research opportunities.

2 Model

We analyze a two-period game between a firm and a worker. Two periods is the minimal time horizon that allows us to capture the impact of effort cost spillovers on the optimal design of the worker's compensation plan and on the firm's performance.

Worker and Firm

For ease of exposition, the worker's output x_t in period $t \in \{1, 2\}$ can take two possible values, with low output ("failure") normalized to 0 and high output ("success") $= x > 0$. In each period t, the worker chooses an effort level $e_t \in E \cup q$, where E is some closed subset of $\mathbb{R}_+$ and q indicates the worker quits. If she does not quit, her effort e_t, which is not observed by the firm, determines the probability of success $p\,(e_t) \equiv Pr(x_t = x|e_t)$ in period t, where $p\,(0) = 0$ and $p\,(e)$ is an increasing function.

Since the firm cannot directly compensate the worker for her effort, it instead motivates her by basing her pay on output, which is mutually observed. Given the binary nature of output, a contract for period t is simply a pair of payout values (f_t, s_t) associated with failure and success in that period. In period t, the worker receives compensation:

$$w_t(x_t) = \begin{cases} f_t & \text{if } x_t = 0 \\ s_t & \text{if } x_t = x \end{cases}$$

We also write $w_t = (f_t, s_t)$. The payments can be thought of as a fixed salary f plus an outcome-dependent bonus $s - f$. Note that the period 2 contract w_2 that the firm offers can depend on the observed period 1 output.

The worker receives increasing utility from income and is assumed to be risk-neutral; we use a simple linear utility function $u(w) = w$. Assuming risk neutrality makes the model more tractable and transparent, allowing us to isolate the effect of spillovers. In a standard model without spillovers, risk aversion leads to firms offering workers reduced incentives and to firms being unable to achieve their first-best outcomes. We obtain both these features with simple risk-neutral workers.

If the worker does not quit, she incurs a cost from the effort she exerts in each period. In period 1, this cost is given by $c_1(e_1)$, where $c_1\,(0) = 0$ and c_1 is strictly increasing with strictly increasing incremental cost,[3] reflecting the common assumption that marginal disutility from effort within a period is increasing. In period 2, the effort cost is given by $c_2(e_1, e_2)$, where i) $c_2\,(0, e_2) = c_1\,(e_2)$[4], ii) c_2 is strictly increasing in e_2 and strictly increasing in e_1 when

[3] When c_1 is differentiable, the incremental cost is simply the marginal cost.

[4] This assumption highlights the role of spillovers when comparing our results to the standard analysis. It

$e_2 > 0$, iii) the incremental cost of second period effort e_2 is strictly increasing, and iv) the incremental cost of second period effort e_2 is strictly increasing in first-period effort e_1. The role of e_1 in $c_2(e_1, e_2)$ reflects the spillover effect from the worker's earlier effort. This spillover could represent a fatigue or burnout effect, as discussed above. In some contexts (e.g., when the worker is a salesperson), it could also represent a saturation effect, in which early-period effort is directed at the easiest tasks or targets – i.e., 'low-hanging fruit' – leaving more difficult ones for the later period.[5]

Special case: A simple special case of spillovers is given by $c_2(e_1, e_2) = (1 + k(e_1)) c_1(e_2) + g(e_1)$, for increasing functions k and g. Spillover k increases the marginal cost of second period effort while spillover g shifts up second period effort cost. We will return to this special case periodically.

- The literature in the standard setting without spillovers often assumes i) $E = \mathbb{R}_+$ ii) p is concave and differentiable, (iii) c_1 is strictly convex and differentiable. Adding, for our context, iv) c_2 is strictly convex in its second argument and differentiable with $\frac{\partial c_2(e_1,e_2)}{\partial e_1 \partial e_2} > 0$, we refer to these four assumptions as the *continuous case*. When E is a finite set, we speak of the *finite case*.[6]

The worker's utility from income and disutility from effort are taken to be additively separable, so that her utility in periods 1 and 2 are given by $w_1(x_1) - c_1(e_1)$ and $w_2(x_2) - c_2(e_1, e_2)$, respectively. The worker maximizes her expected utility. To streamline the analysis, we make the common assumption that when the worker is indifferent between several actions, she chooses the one that is best for the firm.

The worker's best outside option provides utility $\bar{U}$ in any single period, which we normalize to 0. If at any point in time the firm's (anticipated) contracts yield the worker a total expected utility less than 0, the worker rejects the current contract offer, exits the firm, and earns 0 from then on.

can be dropped without affecting our results.

[5] It is also possible for a spillover to be negative, with early effort making later effort *less* costly (perhaps from a learning or 'momentum' effect). Analysis of that case is left for future research.

[6] While E is an unbounded set in the continuous case, this difference with the finite case is inessential. Rather than $E = \mathbb{R}_+$, we can have $E = [0, M]$, with M large enough to assure interior solutions (a sufficient condition is $c(M) > p(M)x$), and everything we will say about the continuous case remains valid.

The firm's profits in each period are defined as output minus the worker's compensation, $\pi_i = x_t - w_t$. (For example, if the worker is a salesperson, then her output would be net sales.) The firm maximizes its expected profits. If at any point the worker exits, the firm earns 0 from then on. Thus, we are assuming that the firm is unable to hire and train a replacement within the time frame of the model.

Sequence of events

The sequence of events is as follows:

1. The firm offers the worker a period 1 contract $w_1 = (f_1, s_1)$.

2. The worker chooses whether to accept the period 1 contract. If she refuses, she quits the firm and the game ends. If she accepts, she chooses her effort level, e_1 and the game continues.

3. Period 1 output x_1 is realized and the firm pays the worker $w_1(x_1)$.

4. The firm offers the worker a period 2 contract $w_2 = (f_2, s_2)$.

5. The worker chooses whether to accept the period 2 contract. If she refuses, the game ends. If she accepts, she chooses her effort level, e_2.

6. Period 2 output x_2 is realized and the firm pays the worker $w_2(x_2)$.

We also consider a variation of this sequence in which the firm commits to contracts for both periods at the outset (in step 1).

3 Analysis and results

Denote the period 1 expected surplus of the worker and firm combined by $E\left[S_1\left(e_1\right)\right] \equiv E[x_1|e_1] - c_1(e_1)$, the period 2 expected surplus by $E\left[S_2\left(e_1, e_2\right)\right] \equiv E[x_2|e_2] - c_2(e_1, e_2)$, and the total two-period expected surplus by $E[S\left(e_1, e_2\right)] \equiv E\left[S_1\left(e_1\right)\right] + E\left[S_2\left(e_1, e_2\right)\right]$. Let e_1^* and e_2^* be the efficient effort levels in periods 1 and 2, which we assume exist and are unique (as is true in the continuous case). Then, $(e_1^*, e_2^*) = \arg\max_{e_1,e_2 \in E \cup q} E\left[S\left(e_1, e_2\right)\right]$. The firm's first-best outcome yields expected profits $E[S\left((e_1^*, e_2^*)\right)$.

It is not readily apparent whether the first-best outcome requires more effort from the worker in the first period or the second. Put another way, is it better for the worker to push herself early then ease up later when effort is more costly, or to hold back early to minimize the cost spillover before making a late push? In fact, both scenarios are possible as we discuss in Section 3.3.

When $e_t^* = 0$, for $t = 1$ or 2, the analysis reduces to a one-period problem, so we restrict our attention to situations in which the efficient outcome requires the worker to exert positive effort in both periods.

- Assumption: $e_1^*, e_2^* > 0$.

Since $e_1^*, e_2^* > 0$, the worker can be profitably employed in both periods. Nevertheless, if her first-period effort is sufficiently high, the spillover cost from that effort may make it *un*profitable to employ her in the second period. In particular, there may exist a first-period effort threshold such that if the worker exceeds it, then she is unwilling to accept *any* contract that the firm is willing to offer in the second period. We call this the "burnout threshold" e_1^b, formally defined by:

$$e_1^b \equiv \min\left\{e_1 \in E : E\left[x_2 \mid e_2\right] - c_2\left(e_1, e_2\right) \leq 0, \forall e_2 \in E\right\}.$$

Suppose e_1^b exists. If $e_1 > e_1^b$, then the firm prefers that the worker exit in period 2. Since the worker exerts positive effort for both periods in the first-best outcome, we have $e_1^* < e_1^b$.

Consider the effort level in E that maximizes period 1 expected surplus $E[S_1\left(e_1\right)]$. We denote this effort by e_1^m, since it is the period 1 *myopic* efficient level. Given any period 1 effort e_1, let $e_2^m\left(e_1\right)$ be the effort level that maximizes period 2 expected surplus $E\left[S_2\left(e_1, e_2\right)\right]$. The following lemma follows (almost) immediately from the negative impact of period 1 effort on period 2 costs.

Lemma 1. *1) $e_1^* \leq e_1^m$, with a strict inequality in the continuous case. 2) $e_2^m\left(e_1\right)$ is weakly decreasing in e_1 - strictly decreasing in the continuous case, whenever $e_2^m > 0$.*

No Spillovers

The standard model without spillovers - $c_2(e_1, e_2) \equiv c_1(e_2)$ - is stationary in character. Hence, whether the model covers a single period or many periods is insignificant, as the well-known single-period analysis extends readily across time. With two periods, efficient effort levels are $e_1^* = e_2^* = e_1^m$. In both periods, the firm captures the maximum possible surplus $E[S_1(e_1^*)]$, while the worker exerts effort e_1^* and receives no surplus. This outcome can be accomplished by the firm offering the contracts $w_t(x_t) = x_t + c_1(e_1^*) - E[x_1|e_1^*], t = 1, 2$. This *franchise*, or sellout, contract can be interpreted as the worker, in each period, keeping the revenue x_t she generates and paying the firm a fixed fee of $E[x_1|e_1^*] - c(e_1^*)$. The firm's first-best outcome obtains whether the players are myopic or long-viewed and whether or not the firm can commit to future contracts. Moreover, details of the model such as whether the worker chooses effort from a continuous or finite set are immaterial.

When there are spillovers, things are quite different. Efficiency is no longer guaranteed and variations of the model can have important impacts, as we will see.

3.1 Myopia

One way that a one-period no-spillover model can be adapted to our setting is to model two separate periods, mechanically adjusting the effort cost in period 2 to account for the spillover from period 1 equilibrium effort. This is equivalent to assuming that players are myopic: In period 1 they both completely disregard their period 2 payoffs. The two periods can then be solved independently in the standard manner, as follows.

In period 1, the firm captures the maximum possible surplus $E[S_1(e_1^m)]$, while the worker exerts effort e_1^m and receives no surplus. This can be accomplished by the firm offering the franchise contract $w_1(x_1) = x_1 + c(e_1^m) - E[x_1|e_1^m]$.

If $e_1^m > e_1^b$ then (by the definition of e_1^b) the firm cannot design a profitable period 2 contract that the worker will accept. Hence, the worker exerts the firm's single-period optimal effort in period 1 but experiences burnout as a result and exits in period 2. If $e_1^m < e_1^b$, the firm can profitably employ the worker in period 2. The firm induces the second-period effort choice $e_2^m(e_1^m)$, which can be done with an appropriate franchise contract. The firm earns period 2 profits $E[S_2(e_1^m, e_2^m(e_1^m))]$, while the worker again derives zero surplus. In both cases, the worker exerts more than the efficient effort level e_1^* in period 1 and less than the efficient level e_2^* in period 2; the firm earns less than its first-best profits $E[S(e_1^*, e_2^*)]$.

Example 1. *Recall the special case $c_2(e_1, e_2) = (1 + k(e_1)) c_1(e_2) + g(e_1)$ and set $c_1(e) = e^2, k(e_1) = ae_1$ and $g(e_1) = be_1$. In addition, let $E = [0, 1]$, $p(e) = e$, and successful output $x = 1$. A simple calculation shows i) $a = 1$, $b = 0 \implies e_1^* = 0.44$, $e_2^* = 0.35$ and $E(S(e_1^*, e_2^*) = .42)$, while ii) $a = 0, b = \frac{3}{5} \implies e_1^* = \frac{1}{5}$, $e_2^* = \frac{1}{2}$ and $E(S(e_1^*, e_2^*)) = \frac{29}{100}$.*[7]

- *In period 1, the myopic firm offers the franchise contract $w_1(x_1) = x_1 - \frac{1}{4}$; the myopic worker chooses period 1 effort $\frac{1}{2} > e_1^*$.*
  - *If $a = 1, b = 0$, the firm offers a period 2 franchise contract $x_2 - \frac{1}{6}$. The worker chooses period 2 effort $\frac{1}{3} < e_2^*$. The firm earns $\frac{5}{12} < E(S(e_1^*, e_2^*)$.*
  - *If $a = 0, b = \frac{3}{5}$, the worker burns out and quits in period 2. The firm earns $\frac{1}{4} < E(S(e_1^*, e_2^*))$.*

Unsurprisingly, a firm suffers when it myopically ignores the cumulative effect of effort on the worker. Generally, however, we expect that firms take a long-term view. Moreover, their experience with workers should provide them with a good understanding of effort spillovers.

On the other hand, workers may be naive and not properly anticipate the draining effect effort has across time. Accordingly, let us suppose that the firm is forward-looking but the worker is myopic. Now the firm can achieve its first-best. In period 1, the firm offers a contract (f_1, s_1) such that $e_1^* = \arg\max_{e \in E}[p(e) s_1 + (1 - p(e)) f_1 - c_1(e)]$. Such a contract must have reduced incentives relative to a franchise contract, $s_1 - f_1 < x$, so that the worker does not over-exert herself. In the above example where $a = 1$ and $b = 0$, in period 1 the firm offers the contract $(f_1, s_1) = (-e_1^{*2}, 2e_1^* - e_1^{*2}) = (-0.19, 0.69)$. Note that $s_1 - f_1 = 0.88 < x = 1$.

We summarize the above findings in the following proposition.

Proposition 1. *When the firm and the worker are both myopic, the firm earns weakly less than its first-best payoff; the worker exerts weakly more than first-best effort in period 1 and*

[7] Note that in i) efficient levels are decreasing while in ii) they are increasing. We return to this in Section 3.3.

weakly less than first-best effort in period 2. All inequalities are strict in the continuous case. When only the worker is myopic, the firm earns its first-best payoff; in period 1 the firm offers the worker a contract with weaker incentives than a franchise contract.

In the remainder of the analysis, we assume that both the firm and the worker are forward-looking.[8]

3.2 Forward-looking firm and worker

We now analyse what might be considered the fully rational model: Both the worker and the firm take full account of spillover effects in a forward-looking manner. For simplicity, we assume that both players maximize the undiscounted sum of their surplus; a small non-zero discount factor would not qualitatively affect the main results. In contrast to the no-spillover case, it is important to consider whether or not the firm can commit to a period 2 contract at the time that it offers a period 1 contract. As we will see, if the firm can commit to a future contract, then it always obtains its first-best outcome; if the firm cannot commit, this is no longer assured. We begin with the commitment case.

Commitment

At the outset of the game, the firm offers a period 1 contract and a period 2 contract. Commitment is not required of the worker, so she is free to quit the firm at any time.

Proposition 2. *When both the firm and the worker are forward-looking and the firm can commit to a future contract, the firm always obtains its first-best outcome. The firm can accomplish this using two franchise contracts. In the continuous case, the worker's expected surplus is negative in period 1 and positive in period 2.*

Let us first consider the properties of a first-best contract in the continuous case. First note that the worker cannot receive a negative expected surplus in the second period, else she would quit after the first period. Suppose her second-period surplus is zero. Then she must be at a myopic best response in period 1; otherwise, she could improve her payoff by

[8] We omit the analysis of the somewhat odd case of a myopic firm and forward-looking worker. In this setting, the myopic firm does not generally obtain its first-best outcome.

myopically optimizing in period 1 and quitting in period 2. Since she is myopically optimizing in period 1, and our functions are differentiable, a slight deviation in her effort level results in essentially no change in her period 1 utility but does affect her period 2 utility, through the spillover effect on her effort cost. The worker can profitably deviate to slightly less effort in period 1 in order to generate surplus utility in period 2 – a contradiction. Therefore, her second-period expected surplus must be positive. At the same time, her first-period expected surplus must be negative since the firm captures the total surplus across both periods. One way for the firm to accomplish this is with the following pair of franchise contracts, which can be used for all parameter values in both the continuous case and the finite case:

$$
\begin{aligned}
w_1(x_1) &= x_1 - E[x_1 \mid e_1^m] + c_1(e_1^m) \\
w_2(x_2) &= x_2 - E[x_2 \mid e_2^*] + c_2(e_1^*, e_2^*) + E[x_1 \mid e_1^m] - c_1(e_1^m) - E[x_1 \mid e_1^*] - c_1(e_1^*)
\end{aligned}
$$

Observe that w_1 here is exactly the same as in the benchmark case, where the firm and worker are both myopic. But now the forward-looking worker does not exert effort e_1^m in period 1. Instead, she internalizes the externality of the spillover and 'shirks', exerting effort $e_1^* < e_1^m$, with a resulting negative payoff of $(E[x_1 \mid e_1^*] - c(e_1^*)) - (E[x_1 \mid e_1^m] + c(e_1^m))$. The firm 'reimburses' her for that loss in period 2.

It is worth noting that this commitment need not be made through a formal future contract offer. In effect, the equilibrium relies on the worker *believing* that the 'correct' contract will be offered in the second period and the firm being sufficiently motivated not to deviate from that offer. For example, although not explicitly represented in our two-period model, that motivation could be the result of the firm's longer-term concerns about its reputation and ability to recruit and retain workers.

The firm's ability to commit to giving the worker a positive surplus in period 2 is crucial. We next consider what happens when the firm is either unable or unwilling to commit to future contracts.

No commitment

Suppose that the firm cannot commit to a period 2 contract in period 1.

Proposition 3. *When both the firm and the worker are forward-looking and the firm cannot commit to a future contract, the firm cannot always obtain its first-best outcome.*

This result stands in contrast to Proposition 2 and to the no-spillover case. To understand Proposition 3, we first show that there is no efficient (subgame perfect) equilibrium in the continuous case. Suppose instead that there is one. As we argued in the commitment case, in an efficient equilibrium the worker must be getting a positive expected surplus in the second period. However, when the firm cannot commit to a future contract, this is impossible. If the worker was obtaining an expected surplus $M_2 > 0$, the firm could modify its second period contract by charging her an additional M_2; the worker would accept the contract without altering her behaviour and the firm would be better off.

This argument can be extended to show that there is actually no subgame perfect equilibrium at all in the continuous case. If the worker is maximizing her period 1 utility, she prefers to reduce her effort (slightly) and benefit from a lower effort cost in any period 2 subgame in which she exerts positive effort. If she is not maximizing period 1 utility, she prefers to maximize her short-term utility and then quit. This seems to suggest a subgame perfect equilibrium in mixed strategies but there is none (see the Appendix). However, this non-existence is essentially a technical issue – recall that existence of a subgame perfect equilibrium is not guaranteed when strategy spaces are continuous. Existence of a (possibly mixed) subgame perfect equilibrium can be guaranteed by using a finite grid to approximate continuous strategy spaces. If the grid is fine enough, the argument of the previous paragraph still applies and all subgame perfect equilibria are inefficient.

While a finite effort choice set has the technical advantage of assuring existence of an equilibrium, there are also good non-technical reasons for its use. The worker may not perceive effort costs in a continuous manner but rather as discontinuous jumps. More to the point, in many contexts the worker's effort choices are more accurately modelled as a finite number of choices. A salesperson's effort, for instance, may be best measured in units like customer calls / visits or days on site. Small changes may have little impact. Or her effort may be predominantly a function of the number of regions or clients she focuses on, so that a set of effort levels with only a few choices, rather than a fine grid with a multitude, may be appropriate. We turn to the finite case next.

Finite case

Suppose the worker chooses among n effort levels, so that $E = (a_1, a_{2,} \ldots, a_n)$, $0 \leq a_1 < a_2 <$ ... $< a_n$. We can think of a_1 as the minimal effort the worker can put in without being fired by her supervisor. If probabilities and costs were defined on an interval of $\mathbf{R}_+$ containing E, the concavity of p and convexity of c would imply $\frac{c_{i+1}-c_i}{p_{i+1}-c_i} \geq \frac{c_i - c_{i-1}}{p_i - c_{i-1}}$ for all i, so we assume these conditions.

Our first proposition covers two extremes: the efficient first-period effort is very low, $e_1^* = a_1$, or very high, $e_1^* = e_1^m$. The former case implies that spillover costs are important; the latter may correspond to negligible spillovers or to non-negligible spillovers which reveal themselves in a small e_2^*.

Proposition 4. *If $e_1^* \in \{a_1, e_1^m\}$, the firm obtains its first-best outcome. When $e_1^* = a_1$, the firm can accomplish this using a flat contract in period 1 followed by a franchise contract in period 2. When $e_1^* = e_1^m$, the firm can offer two franchise contracts.*

First suppose that $e_1^* = a_1$. If the firm offers a flat contract in the first period, the worker responds by putting in minimal effort a_1; a franchise contract in the second period elicits efficient behaviour in the usual way. In each period, the firm extracts all expected surplus. Now suppose that $e_1^* = e_1^m$ and that the firm offers two franchise contracts. As a general matter, such an offer either i) elicits efficient behaviour or ii) induces the worker to myopically over-exert herself in period 1 and then quit. When $e_1^* = e_1^m$, myopic optimization is efficient so that ii) does not apply.

Our next two propositions cover spillovers that reduce the first-period efficient output, though not to an extreme, $a_1 < e_1^* < e_1^m$. The first proposition gives sufficient conditions for the firm to achieve its first-best outcome.

Consider a first-best equilibrium. The worker exerts efforts e_1^*, e_2^* and earns a two-period (expected) surplus of 0. Since there is no contract commitment, the firm will extract all the surplus in period 2. Hence, in period 1 the firm must offer a contract under which e_1^* is the worker's myopic best response; otherwise the worker could do better by myopically optimizing in period 1 and quitting in period 2. Note that the worker earns no surplus in period 1 either, since her total surplus across periods is 0.

Let $e_1^* = a_g$ and $e_2^* = a_h$ and define the first period contract $w_1 = (f^*, s^*)$ by

$$f^* = \frac{p(a_{g+1})\, c_1(a_g) - c_1(a_{g+1})\, p(a_g)}{p(a_{g+1}) - p(a_g)},\ s^* = \frac{(1 - p(a_g))\, c_1(a_{g+1}) - c_1(a_g)(1 - p(a_{g+1}))}{p(a_{g+1}) - p(a_g)} \quad (1)$$

This contract is designed to satisfy the conditions described in the previous paragraph, which are similar to those found in a standard principal-agent problem without spillovers. First, the contract satisfies the incentive compatibility constraint, ensuring that $e_1^* = a_g$ is the worker's myopic optimal effort choice. Second, the worker's participation constraint binds at that effort, ensuring that the firm collects all period 1 expected surplus.

Under this contract, the worker cannot benefit *in the short term* by deviating from e_1^*. However, due to her effort cost spillover she might still prefer to reduce her effort in period 1 to enjoy a lower effort cost in period 2. For the firm to achieve its first-best outcome, the firm must be able to ensure that the worker's loss from any shirking in period 1 outweighs her increased efficiency in period 2. The condition in Proposition 5 guarantees this.

Proposition 5. *Suppose $e_1^* \notin \{a_1, e_1^m\}$. The firm can obtain its first-best outcome if, for all $i \le g$ and $j \ge h$,*

$$p(a_i)\, s^* + (1 - p(a_i))\, f^* - c_1(a_i) \le p(a_h)\, x - c_2(a_g, a_h) - p(a_j)\, x + c_2(a_i, a_j)\,. \quad (2)$$

In period 1, the firm must offer incentives that are stronger than a flat contract, to induce $e_1 > a_1$, but weaker than a franchise, to preclude $e_1 = e_1^m$. In the second period, the firm can offer a franchise contract.

The contract $w_1 = (f^*, s^*)$ satisfies the incentive compatibility constraint, while $s^* - f^*$ is the largest incentive that the firm can offer without inducing a first-period effort choice greater than e_1^*. That is, under the contract $w_1 = (f^*, s^*)$, effort e_1^* is optimal in period 1, while under any contract (f_1, s_1) with $s_1 - f_1 > s^* - f$, some effort $e_1 > e_1^*$ yields the worker a higher period 1 payoff. Thus, the contract provides the strongest possible disincentive for the worker to shirk in period 1. Condition (2) ensures that when the firm offers (f^*, s^*) in period 1 followed by a franchise contract in period 2, the worker's first-period loss from choosing a smaller effort than e_1^* outweighs her resulting second-period gain from reduced effort costs.

When condition (2) is violated, it is not always possible for the firm to achieve its first-best outcome. As the next proposition indicates, the firm's best strategy might then be to burn the worker out – maximizing her short term effort then letting her exit.

Proposition 6. *Suppose $e_1^* \notin \{a_1, e_1^m\}$. When condition (2) does not hold, the firm may not be able to achieve its first-best outcome. When it cannot achieve its first-best, the firm sometimes offers a burnout contract in period 1. The resulting inefficiency can be severe: The firm's expected payoff under a burnout equilibrium can approach half of the first-best, but can never be lower.*

Although the proposition focuses on burnout, there may also be other types of inefficient equilibria when condition (2) fails. One possibility is an equilibrium in which the worker mixes between e_1^* and some lower action $a_i < e_1^*$. In such an equilibrium, the firm offers a contract with a higher franchise fee in period 2 when output is low in period 1.

The exact conditions under which burnout is optimal for the firm may be affected by additional factors that are not included in our model. For example, we have assumed that workers who exit cannot be replaced in period 2. It is possible that the firm could hire and deploy a new worker quickly enough to achieve some productivity in period 2. In that case, burning workers out would become more attractive, effectively relaxing the conditions under which such an equilibrium exists. On the other hand, we have not included some additional costs associated with turnover (e.g., hiring and training new workers), which would make burnout less attractive and thus narrow those conditions. Even when it is the firm's optimal strategy, burning the worker out can be very costly relative to the first-best outcome, as illustrated in Example 3.

While we have assumed that the worker is risk-neutral in order to isolate the effects of effort cost spillovers, our main qualitative results, including the firm's inability to achieve its first-best outcome and the existence of burnout equilibria, will continue to hold with some risk aversion. We leave the complete analysis of a risk-averse agent for future research

Two Examples In this section, we present two examples that illustrate propositions 5 and 6. We allow the worker's effort to take one of three values - $E = (a_1, a_2, a_3)$ - which is the simplest finite case that displays a range of outcomes. In both examples, the probability of success and first-period costs are described by the following matrix:

$$\begin{array}{ccc} \text{Period 1 effort} & p\left(a_i\right) & c_1\left(a_i\right) \\ a_1 & \frac{1}{100} & 1 \\ a_2 & \frac{52}{100} & 2 \\ a_3 & \frac{58}{100} & 4 \end{array} \tag{3}$$

The first example illustrates Proposition 5 - the sufficient condition (2) is satisfied, so the firm can obtain its first-best outcome.

Example 2. *The probability of success and the first-period effort costs are given by matrix* (3)*. Second-period effort costs* $c_2\left(a_i, a_j\right)$ *are given by the matrix:*

Period 2 effort $\rightarrow$	a_1	a_2	a_3
Period 1 effort $\downarrow$	$c_2\left(a_i, a_1\right)$	$c_2\left(a_i, a_2\right)$	$c_2\left(a_i, a_3\right)$
a_1	1	2	51
a_2	2	17	72
a_3	300	450	500

It is easily verified that the efficient effort choices are $e_1^* = a_2$, $e_2^* = a_2$*. Moreover, these are also the unique equilibrium effort levels. These efforts can be implemented by the contract* $w_1 = \left(f_1^*, s_1^*\right) = \left(-\frac{46}{3}, \frac{54}{3}\right)$*, as defined by (1), followed by the contract* $w_2 = \left(f_2, s_2\right) = \left(-9, 41\right)$*. The firm earns the first-best surplus* 813*. (See Appendix for details).*

The second example illustrates Proposition 6; condition (2) is not satisfied and the firm cannot obtain its first-best outcome. It differs from the first example in that the cost spillover from period 1 effort has increased. For simplicity, the only change is an increase in the cost of effort level a_2 in period 2 following a_2 in period 1.

Example 3. *The probability of success and the first-period effort costs are given by matrix*

(3). *Second-period effort costs* $c_2(a_i, a_j)$ *are given by the matrix:*

Period 2 effort $\rightarrow$	a_1	a_2	a_3
Period 1 effort $\downarrow$	$c_2(a_i, a_1)$	$c_2(a_i, a_2)$	$c_2(a_i, a_3)$
a_1	1	2	51
a_2	2	22	72
a_3	300	450	500

Again, $e_1^* = a_2$, $e_2^* = a_2$. *But now these effort levels cannot be achieved in equilibrium. Instead, the unique equilibrium outcome is an effort of* a_3 *in period 1, resulting in the worker burning out and quitting. This outcome can be implemented by the contract offers* $w_1 = (f_1^*, s_1^*)$ *and* $w_2 = (0, 0)$. *The firm earns a surplus of* 460, *which is* 57% *of the first-best surplus* 808.

Why is it that efficiency no longer obtains? The firm could, as before, offer a contract that makes the effort level a_2 *myopically optimal in period 1,*[9] *followed by a franchise contract. But now, the increased spillover cost means that the (forward-looking) worker would choose effort level* a_1 *in period 1, sacrificing some of her initial payoff in favour of a reduction in her later effort cost.*

Note that equilibrium effort levels are constant in the first example and decreasing in the second. It is also possible for them to be increasing. For instance, this occurs when $a_1 = e_1^*$ and efficient efforts are increasing.

In the next section, we consider the direction of efficient effort levels.

3.3 The Effect of Spillovers on First-Best Efforts

With no spillovers, efficient effort levels are a constant e^* across periods. Spillovers (weakly) reduce the efficient levels. Which period's effort level is affected more? Should the worker work relatively hard in the first period and ease up later when effort is more costly, or should she cut back sharply in the first period to minimize the magnitude of the spillover? To gain some insight into these questions, we recall the special case $c_2(e_1, e_2) = (1 + k(e_1))\, c_1(e_2) +$

[9] In fact, with the contract (f^*, s^*) both a_2 and a_3 are myopically optimal (this indifference is inconsequential and could be broken with a slight adjustment.)

$g(e_1)$ and suppose that $E = \mathbb{R}_+$, g is differentiable, and $c_1(e_1) = be_1^y$ and $k(e_1) = ae_1^z$. Thus,

$$c_1(e_1) = be_1^y,\ c_2(e_1, e_2) = (1 + ae_1^z)\, be_2^y + g(e_1)$$

Our maintained assumptions imply that $y > 1$ and $z, a, b, g'(x) \geq 0$, with $a = 0 \implies g' > 0$ and $g' \equiv 0 \implies a > 0$.

Proposition 7. *1. If $a = 0$, efficient effort levels are increasing.*

2. If $z > y$, efficient effort levels are increasing.

3. If $z < y$ and $g' \equiv 0$, efficient effort levels are decreasing.

4. If $z = y$ and $g' \equiv 0$, efficient effort levels are constant.

When $a = 0$, the spillover does not affect the marginal cost of second-period effort. As a result, the second-period efficient level is the same as without spillovers. On the other hand, the first-period efficient level is reduced due to the negative externality reflected in $g'(x) > 0$, so that $e_1^* < e^* = e_2^*$ (see Example 1). Now, suppose $a > 0$. When $z > y$, the spillover cost from effort is more convex than the direct cost of effort. This leads to a large reduction in first-period effort to mitigate the second-period externality, and again $e_1^* < e_2^*$. When $z < y$, the spillover cost is less convex than the direct cost. In and of itself, this lessens the impetus for a reduction in first-period cost from the stationary optimum e^*, suggesting that $e_1^* > e_2^*$. However, this pattern could be reversed by the effect of an increasing fixed cost, hence 3) adds the condition $g' \equiv 0$ for an unambiguous result (see Example 1). While the intuition behind these results is reasonable enough, we cannot at this point claim that the findings extend to more general cost functions.

When $a > 0$ and $g' > 0$, we have $e_1^*, e_2^* < e^*$. Although this might suggest that efficient levels are decreasing in spillover costs, the effect is not so simple. For instance, suppose that $g(e_1) = re_1$. As r increases, e_1^* does fall. But this fall in e_1^* decreases second-period marginal costs, so that e_2^* rises. On the other hand, an increase in a can, in some instances, cause a rise in e_1^* and a fall in e_2^*.

4 Conclusions

Employee burnout is a significant issue that has long plagued firms, gaining increased attention in recent years. The prevalence of burnout is an indication that the costs of work-related effort (such as fatigue) are incurred by workers not only *while* they are working, but for some time after. We incorporate this effort cost spillover into a two-period principal-agent model to study its impact on the optimal design of incentive contracts over time.

The single-period principal-agent model typically used to study static incentive contracts under moral hazard cannot account for these dynamic spillover effects. Applying the standard franchise solution in the presence of spillovers results in the employee exerting more than the efficient effort in the first period and less in the second; the firm obtains less than first-best profits.

Using a dynamic model, we find that a forward-looking firm employing a myopic worker *can* achieve its first-best profit. It does so by offering weaker incentives in the first period than it would without spillover costs, thus inducing the worker to optimally reduce her first period effort.

When the firm and the worker are both forward-looking the analysis is trickier. If the firm can commit to a multiperiod contract, it can always obtain first-best profits. Otherwise, it may not be able to. In some cases, the firm's best possible outcome is a 'burnout equilibrium', in which the worker exerts so much effort in period 1 that she quits in the second period. This can be the case even when retaining the worker over time would generate almost twice as much surplus as burning them out.

These results have a number of managerial implications. In order to maximize profits over time, firms should consider the long-term effects of effort on employees' well-being and willingness to work and adjust expectations and incentives accordingly. The existence of burnout equilibria might help to explain the prevalence of employee burnout, even when firms can seemingly benefit by increasing retention. For firms seeking to reduce burnout, our results suggest that they might do so by committing to longer-term incentive contracts. Similarly, a policymaker seeking to decrease burnout rates should consider interventions that encourage or enable firms to commit to longer-term contracts. Finally, our results indicate the value to firms of reducing effort cost spillovers among their employees. This might help to explain the growing popularity of benefits such as employee wellness programs and flexible work arrangements, among others.

We conclude with a brief discussion of some of our model assumptions and limitations, which suggest opportunities for further research. We have made the simplifying assumption that a worker's outside option utility in the second period is unaffected by her first-period effort. However, if her effort cost spillover represents fatigue, then she might continue to feel that effect if she leaves and joins another firm, thereby lowering her outside option utility. Conversely, if the spillover reflects task saturation, then instead of affecting the worker's outside option, it might continue to affect the firm (or the worker's replacement) after she exits. It would be interesting, both theoretically and managerially, to consider how our results would be affected if the spillover cost of effort was subject to some exogenous shock, forcing the firm to design the second-period contract with imperfect information about the worker's effort cost. Finally, this research is motivated by common concerns about employee burnout, so we have focused on positive effort cost spillover, with the worker's period 2 effort cost increasing in period 1 effort. It might be worth considering the effects of negative spillovers which could exist if early effort generates some type of learning or momentum effect.

References

Ábrahám, A., & Pavoni, N. (2008). Efficient allocations with moral hazard and hidden borrowing and lending: A recursive formulation. *Review of Economic Dynamics*, *11*(4), 781–803.

Banker, R. D., Lee, S. Y., Potter, G., & D., S. (2001). An empirical analysis of continuing improvements following the implementation of a performance-based compensation plan. *Journal of Accounting and Economics*, *30*(3), 315–350.

Basu, A. K., Lal, R., Srinivasan, V., & Staelin, R. (1985). Salesforce compensation plans: An agency theoretic perspective. *Marketing Science*, *4*(4), 267–291.

Bouwens, J., & Kroos, P. (2011). Target ratcheting and effort reduction. *Journal of Accounting and Economics*, *51*(1-2), 171–185.

Cordes, C. L., & Dougherty, T. W. (1993). A review and an integration of research on job burnout. *Academy of Management Review*, *18*(4), 621–656.

Cordes, C. L., Dougherty, T. W., & Blum, M. (1997). Patterns of burnout among managers and professionals: A comparison of models. *Journal of Organizational Behavior*, *18*(6), 685–701.

Dearden, J. A., & Lilien, G. L. (1990). On optimal salesforce compensation in the presence of production learning effects. *International Journal of Research in Marketing*, *7*(2-3), 179–188.

Demarzo, P. M., & Sannikov, Y. (2017). Learning, termination, and payout policy in dynamic incentive contracts. *The Review of Economic Studies*, *84*(1), 182–236.

Fudenberg, D., Holmström, B., & Milgrom, P. (1990). Short-term contracts and long-term agency relationships. *Journal of Economic Theory*, *51*, 1–31.

Gershkov, A., & Perry, M. (2012). Dynamic contracts with moral hazard and adverse selection. *Review of Economic Studies*, *79*(1), 268–306.

Godes, D. (2004). Contracting under endogenous risk. *Quantitative Marketing and Economics*, *2*(4), 321–345.

Habel, J., Alavi, S., & Linsenmayer, K. (2021). Variable compensation and salesperson health. *Journal of Marketing*, *85*(3), 130–149.

Holmström, B. (1979). Moral hazard and observability. *The Bell Journal of Economics*, *10*(1), 74–91.

Holmström, B., & Milgrom, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, *55*(2), 303–328.

Indjejikian, R. J., Matějka, M., & Schloetzer, J. D. (2014). Target ratcheting and incentives: Theory, evidence, and new opportunities. *The Accounting Review*, *89*(4), 1259–1267.

Jerath, K., & Long, F. (2020). Multiperiod contracting and salesperson effort profiles: The optimality of "hockey stick," "giving up," and "resting on laurels". *Journal of Marketing Research*, *57*(2), 211–235.

Joseph, K., & Thevaranjan, A. (1998). Monitoring and incentives in sales organizations: An agency-theoretic perspective. *Marketing Science*, *17*(2), 107–123.

Kronos (2017). The employee burnout crisis: Study reveals big workplace challenge in 2017. `https://www.kronos.com/about-us/newsroom/employee-burnout-crisis-study-reveals-big-workplace-challenge-2017`.

Laffont, J. J., & Tirole, J. (1988). The dynamics of incentive contracts. *Econometrica*, *56*(5), 1153–1175.

Lal, R., & Srinivasan, V. (1993). Compensation plans for single- and multi-product salesforces: An application of the Holmstrom-Milgrom model. *Management Science*, *39*(7), 777–793.

Lal, R., & Staelin, R. (1986). Salesforce compensation plans in environments with asymmetric information. *Marketing Science*, *5*(3), 179–198.

Lambert, R. A. (1983). Long-term contracts and moral hazard. *The Bell Journal of Economics*, *14*(2), 441–452.

Mantrala, M. K., Raman, K., & Desiraju, R. (1997). Sales quota plans: Mechanisms for adaptive learning. *Marketing Letters*, *8*(4), 393–405.

Maslach, C., Schaufeli, W. B., & Leiter, M. P. (2001). Job burnout. *Annual Review of Psychology*, *52*(1), 397–422.

Rogerson, W. P. (1985). Repeated moral hazard. *Econometrica*, *53*(1), 69–76.

Schaufeli, W. B., & Bakker, A. B. (2004). Job demands, job resources, and their relationship with burnout and engagement: A multi-sample study. *Journal of Organizational Behavior*, *25*(3), 293–315.

Schöttner, A. (2017). Optimal sales force compensation in dynamic settings: Commissions vs. bonuses. *Management Science*, *63*(5), 1529–1544.

Steenburgh, T. J. (2008). Effort or timing: The effect of lump-sum bonuses 6(3), 235. *Quantitative Marketing and Economics*, *6*(3), 235–256.

Sullivan, T. (2018). How to stretch your team without straining them. https://www.peoplemanagement.co.uk/voices/comment/stretch-your-team-without-straining.

Weitzman, M. L. (1980). The "ratchet principle" and performance incentives. *The Bell Journal of Economics*, *11*(1), 302–308.

World Health Organization (2018). International classification of diseases for mortality and morbidity statistics (11th revision).

Appendix

Proofs and detailed solutions for examples to be added.