# Hidden Overtime: Optimal Contracts with (Self-)Deceptive Effort Reports *

Juho Alasalmi[†] University of Konstanz
Susanne Goldlücke[‡] University of Konstanz
Michelle Jordan[§] University of Konstanz

April 2023

**Abstract**

Requiring extra hours to get a job done signals low skills. With unobservable effort (hours of work), image-concerned agents may prefer to underreport effort to hide low skills from the principal or themselves. We show how such "hidden overtime" can arise as a consequence of the optimal contract if the principal asks for overtime reports but has no way of ensuring that these reports are also correct. It is possible that the principal benefits from the agent's image concerns but also that the agent works inefficiently long hours. While a recording system that makes underreporting difficult can alleviate the inefficiency caused by hidden overtime, a legal obligation to install such a system is not necessary.

*JEL classification:* D91, D82, D86, K31
*Keywords:* self-signaling, hidden effort, career concerns

## 1 Introduction

Advances in information and communication technologies and the effects of the global COVID-19 pandemic have led to a change in the nature of work. Working from home or rented office spaces has become more and more common. Telework exacerbates monitoring problems, which makes employers fear a decrease in workers' effort. At the same time, there is a growing concern that as working time gets more difficult to measure, workers may put in longer hours. While this phenomenon of "hidden overtime" is not new,[1] it has received increased attention with the spread of teleworking. For example, the International Labour Organization reports that teleworkers tend to work longer hours and more unpaid hours than workers working in the premises of their employers (ILO, 2017). Some authors even speak of an "epidemic of hidden overtime" and call for additional legal regulation of working hours (Jones et al., 2021).

*We thank Andreas Asseyer, Jean-Michel Benkert, Urs Fischbacher, Tim Friehe, Mats Godenhielm, Jeanne Hagenbach, Max Lobeck, Marc Möller, Ferdinand von Siemens, John Douglas Skåtun, Irenäus Wolff and many other participants at workshops and seminars for valuable comments and discussions.

[†]University of Konstanz, Department of Economics, Box 150, 78457 Konstanz, Germany. Email: juho.alasalmi@uni-konstanz.de. Phone: +49 7531 88 3502

[‡]University of Konstanz, Department of Economics, Box 150, 78457 Konstanz, Germany. Email: susanne.goldluecke@uni-konstanz.de. Phone: +49 7531 88 2989

[§]michelle.jordan.mail@gmail.com

[1]For instance, in the surveys of Smith et al. (1995) 89% of tax professionals, Otley and Pierce (1996) 54% of audit seniors, Akers et al. (1999): 71% of accountants, Barrainkua and Espinosa-Pike (2015) 55% of audit partners, Coulshed et al. (2022) 84% of doctors-in-training admitted having underreported their working hours.

In many countries, employers already have an obligation to care for the well-being of their employees. This has recently led the European Court of Justice to determine that employers also have an obligation to record the actual working time of their employees (ECJ, C55/18, 2019). However, given the delegated nature of work and the implausibility of perfect monitoring, it is not clear whether such an obligation will succeed in overcoming the problem of hidden overtime, and whether there might be unwanted side effects.

To study the potential effects of such regulation, it is important to understand the determinants of hidden overtime. Supplemental work from home is often motivated by "catching up on work" (ILO, 2017, page 22). Underreporting of hours has been documented to correlate with feelings of incompetence (Barrainkua and Espinosa-Pike, 2015) and time pressure (Barrainkua and Espinosa-Pike, 2015; Otley and Pierce, 1996) as well as with perceptions that underreporting improves performance evaluations and career prospects, increases job satisfaction and leads to feeling better about oneself (Akers and Eaton, 2003; Coulshed et al., 2022). In light of these findings, we suggest that hiding overtime is about signaling ability and build a theoretical model that includes completion time, image or career concerns, and monitoring.

We study a labor relationship in which a principal contracts with an agent who wants to be perceived as high-skilled, by the principal or by himself. The contract can specify a transfer based on task output and the agent's reported working time. The agent completes the task only if he either has high skills or does costly overtime work. The principal does not observe the agent's skill nor the actual working hours. By working unreported overtime hours the agent can thus hide low skills from the principal or even from himself if he does not perfectly recall how long it took him to complete the task. Concealing overtime work may reduce the agent's payment, but results in a more competent (self-)image.

We explore the consequences of image concerns for the principal's optimal contract. Without image concerns, the agency problems in this model may lead to an under-provision of effort. The principal therefore potentially benefits from image concerns, which mitigate the moral hazard problem. This is indeed the case if the benefit of a completed task exceeds the cost of working overtime. If image concerns are large but working overtime is not efficient, then the image concerns may backfire and the agent works inefficiently long hours. In both these cases, there is hidden overtime, but only in the second case, there would be an efficiency gain if overtime was revealed instead.

We then examine potential channels through which the obligation to install a recording system can have an effect. The benefits of a recording system depend on how easy it is to manipulate the system. If the only possibility is to collect cheap talk reports of employees, who for example work from home or are impossible to monitor for other reasons, then an obligation to record hours worked has no effect apart from the recording cost since the agent may (self-)deceive with his reports. In contrast, if the agent has to be present at the facility or use a technical system in order to work, then it should be feasible to prevent the agent from understating hours worked. For example, the employer could make the agent log on to a time recording device upon arrival and enforce lunch breaks. Then understating working hours is infeasible while overstating working hours, for example by taking unreported breaks, remains possible. We show that such a system always at least weakly increases welfare. However, we also argue that the principal already has sufficient incentives to install such

a system. Monitoring systems that deter overstating hours worked, on the other hand, may either increase or decrease the agent's rent and the principal may have insufficient or excessive incentives to install such systems. However, giving the agent some power in this decision already leads to an efficient bargaining solution.

Following a discussion of related literature in Section 2, the model is presented in Section 3. In Section 4, we study the behavior of an agent given a fixed contract and derive the optimal contracts for different monitoring regimes. In Section 5, the consequences for regulation are discussed. Section 6 concludes. All proofs are relegated to the appendix.

# 2 Literature

Hiding overtime seems on one hand to be motivated by performance evaluations and career prospects but on the other hand by feeling good about oneself and protecting one's self-esteem. We thus motivate agent's image concerns as either stemming from the perceptions of the principal or the agent himself and relate our model to the literatures on career concerns and self-deception.

## 2.1 Career Concerns

With hidden effort and output potentially informative about hidden skills, our model relates to the career concerns models (e.g. Holmström (1999); Gibbons and Murphy (1992)). In these models, output is a stochastic function of effort and skill. Agent's future wages are determined by the perceptions that the principal (or market) forms about agent's skill after observing his output. As output is increasing in skill the inferred skill is increasing in output. The agent's effort provision is motivated by future wages and thus by the aim to make the inference about his skills more favorable as the agent and principal (or market) symmetrically learn about the agent's skills.

When effort is observable and information about employees' skills asymmetric, effort becomes a signal. This insight goes back at least to Spence (1973), who models education as a signal of skills to the labor market. Effort on the job, if observed, can work as a signal as well: Sousa-Poza and Ziegler (2003) and Anger (2008) model high effort on the job as a signal of high productivity by supposing, in line with Spence (1973), that productivity and disutility from work correlate negatively. In Akerlof (1976), workers in assembly lines of different speeds are compensated by the average output in their assembly line. This creates an incentive to pool with harder-working colleagues in faster assembly lines. Landers et al. (1996) model high effort as a signal of low disutility of effort: Working long hours improves labor market prospects since hard-working colleagues are desirable joint project partners.

While in this literature workers wish to signal their employer that they are willing to put in long hours, we suggest that the workers may wish to signal high productivity by hiding the long hours that it took to produce a given output. In Sampson (2002), long hours may, in addition to signaling low disutility of effort, signal low productivity. While in Sampson (2002) effort is observable, Aron (1987) and Kőszegi and Li (2008), as we, combine unobservable effort and asymmetry of information in a career concerns model. The agent has private information

about his disutility of effort (or so-called "drive" in Kőszegi and Li (2008)) and may not always want to reveal it to the principal since while appearing "driven" is initially rewarded it may also, for given output, signal low skills. In Aron (1987), the principal might exploit such information in the second period, and therefore the optimal screening contract might let the agent types pool in the first period.

### 2.2 Self-Deception

The willingness to take actions to retain positive self-image in the sense of a belief in high ability may arise from multiple sources: High self-image may have instrumental value as a commitment device in future wage bargaining (Hvide, 2002; Bénabou and Tirole, 2009) or as a boost to motivation to counteract present bias (Bénabou and Tirole, 2002, 2004). High self-image may have signaling value in reducing signaling costs when trying to convince others about high ability (Trivers, 2011; Schwardmann and van der Weele, 2019). Finally, self-image may simply have consumption value (Schelling, 1986) as a direct value of self-esteem (Köszegi, 2006) or, insofar as high ability is rewarded with high income, as anticipation of high future consumption (Löwenstein, 1987).

The ability to retain positive self-image by taking certain actions requires denial or imperfect recall. Such self-deception has been studied in experimental and theoretical cognitive psychology. In the experiment of Quattrone and Tversky (1986), subjects who were told that a high tolerance of cold water is indicative of good future health, tolerated cold water longer than subjects who were told the opposite. The majority of subjects denied that the cover story influenced their tolerance and increased their optimism about their future health. Those who admitted having tried to manipulate their tolerance to cold water were not as optimistic about their future health as others. This sort of effort denial has later been replicated in the experiments of Sloman et al. (2010) and Fernbach et al. (2014). Fernbach et al. (2014, page 6) summarize "As people change their behavior to provide positive evidence for a desirable trait, they simultaneously deny doing so in order to enhance the diagnosticity of the evidence for the positive trait."

Self-deception thus seems to require simultaneous forgetting of the task difficulty and the effort exerted. Later ambiguity about the task difficulty facilitates the malleability of beliefs about productivity and underestimation of exerted effort supports beliefs of high productivity. Self-deception is successful due to the uncertainty of whether the observable output is the result of task difficulty or effort provision. Reffett et al. (2014) experimentally study underreporting of working time and find that it is the time between exerting effort and reporting effort that facilitates underreporting pointing to the important role of memory.[2]

Effort denial distinguishes our model from the game-theoretic models of self-signaling (Bénabou and Tirole, 2004; Bernheim and Thomadsen, 2005; Bénabou and Tirole, 2006; Mijović-Prelec and Prelec, 2010; Bénabou and Tirole, 2011). In these models, actions are

[2]The role of memory for self-signaling is also studied by van der Weele and von Siemens (2020), who ask whether subjects who received unfavorable feedback are willing to pay to more easily forget this information. They find that people instead adjust the importance they assign to self-image. Consistent with this finding, in our model the importance of self-image and propensity to forget are substitutes. However, the importance of self-image is fixed.

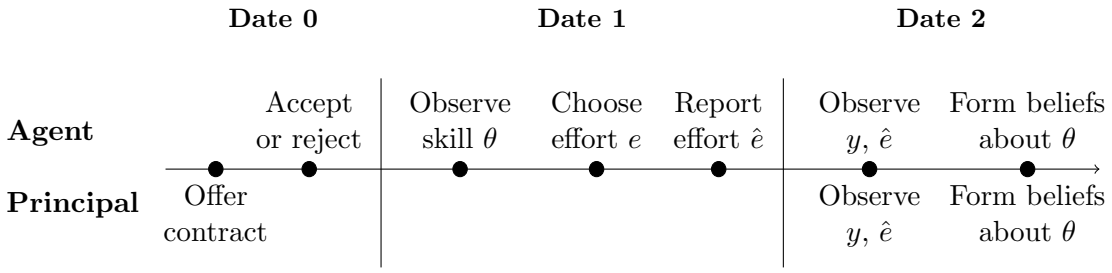


Figure 1: Timeline.

signals to future selves and denial is with respect to an underlying trait. In our model, the signal is an outcome jointly caused by action and an underlying trait and denial is with respect to an action and an underlying trait.[3] Self-deception is facilitated by the denial of the action that produces a favorable signal. Thus, our modelling of the agent's behavior can be interpreted as a model of self-deception with effort denial. We thus contribute to the formal modelling of self-signaling and self-deception with a mechanism that closely corresponds to the literature on self-deception in cognitive psychology.

# 3 The Model

We study a principal-agent relationship with hidden action and hidden information. The agent (he) can be interpreted as an employee or a self-employed or a freelance worker and the principal (she) as an employer or a customer, respectively. Both agent and principal are risk-neutral. The agent has image concerns and is wealth-constrained such that payments to the agent must be non-negative.

## 3.1 The Sequence of Events

The timeline is illustrated in Figure 1. At date 0, the principal offers a contract which the agent either accepts or rejects. If the agent rejects, both receive zero as their pecuniary outside option payoffs.

If the agent accepts the contract, the agent is confronted with a task and only then privately learns his skill or productivity in the task, denoted by $\theta \in \{\theta_L, \theta_H\}$ with $0 < \theta_L < \theta_H$. Interpreting $\theta$ as task-specific productivity, it is natural to assume that it is not known in advance how difficult the task will be, but that this becomes evident to the agent once he works on the task. The agent is high-skilled, $\theta = \theta_H$, with probability $p \in (0,1)$ and low-skilled, $\theta = \theta_L$, with probability $1-p$. Expected productivity is denoted by $\bar{\theta} = p\theta_H + (1-p)\theta_L$ and the range of productivity by $\Delta = \theta_H - \theta_L$. While the principal never observes $\theta$, the type distribution is common knowledge.

After having privately observed his skill $\theta$, the agent chooses effort $e \in \{e_L, e_H\}$ at cost

[3] Bénabou and Tirole (2004) and Bénabou and Tirole (2011) allow the signaling actions to be forgotten with exogenous probability. In their models, forgetting an action thus prevents self-signaling whereas in our model forgetting an action facilitates self-signaling.

$c(e)$. The cost of low effort is normalized to zero, while high effort has a cost of $c > 0$:

$$c(e) = \begin{cases} c & \text{if } e = e_H \\ 0 & \text{otherwise} \end{cases}. \tag{1}$$

We interpret effort as working hours: $e = e_L$ as "working the regular working hours" and $e = e_H$ as "working overtime".

Both the agent's skill and the exerted effort contribute to output $y \in \{y_L, y_H\}$, where $y = y_H$ means that the task is successfully completed, while $y = y_L$ means that the task was not successfully completed. Binary output is simple but encompasses a large range of situations. Many tasks, like repairing a product or writing code for a certain purpose, have a pass-fail-structure. Effort and skill map deterministically into output according to the following production function:

$$y(\theta, e) = \begin{cases} y_L & \text{if } \theta = \theta_L \wedge e = e_L \\ y_H & \text{otherwise} \end{cases}. \tag{2}$$

A high-skilled agent can thus complete the task during regular hours, whereas a low-skilled agent completes the task if and only if he works overtime. Put differently, $y_H$ is defined as the output that a high-skilled agent can get done in regular time, $y_L$ is the output of a low-skilled agent in the same time, and overtime is the time that the low-skilled agent needs to catch up on the high-skilled agent.

### 3.2 Contracts and Effort Reports

Contracts cannot condition on the agent's effort or task difficulty. Output is observable and can be contracted upon. Also, the contract can condition on the agent's effort report. Time cards, time sheets, time books and time-tracking software are all ways for the employer to ask the employee how long he worked. Self-employed and freelance workers may bill their customers by the time it took to complete the work.

Our analysis will focus on different regimes regarding the manipulability of such time recording systems and the legal obligation to install them. We achieve this flexibility in the model by introducing a cost of misrepresenting effort in the recording system. This cost may be a psychic lying cost or an effort cost of manipulating a monitoring system. Thus, we assume that the agent incurs a cost $l_o \geq 0$ if he reports $\hat{e} = e_H$ when he only worked $e = e_L$, and a cost of $l_u \geq 0$ if he reports $\hat{e} = e_L$ when he worked $e = e_H$.

The output $y$ and the agent's self-reported effort $\hat{e} \in \{e_L, e_H\}$ jointly determine the transfer $W(y, \hat{e})$ that the agent receives. The agent faces a limited liability constraint, which means that for all $y, \hat{e}$:

$$W(y, \hat{e}) \geq 0. \tag{3}$$

Another interpretation of this constraint is that there is a minimum wage that is normalized to zero.

We will focus on a class of contracts which stipulate a bonus $b \geq 0$ for a successfully

completed task and an overtime payment $w \geq 0$ which depends on the agent's report. The overtime payment is paid if the agent completes the task and reports overtime work. Hence, we make the following additional restrictions on contracts:

$$W(y_L, \hat{e}) = 0, w = W(y_H, e_H) - W(y_H, e_L) \geq 0 \text{ and } b = W(y_H, \hat{e}) - W(y_L, \hat{e}) \geq 0.^4$$

### 3.3 Information and Payoffs

In addition to transfers, the agent is motivated by image concerns. As discussed in Section 2, the value of image may stem from various sources. We consider here the interpretation that the agent directly cares about self-image or about how he is viewed by the principal (or his coworkers), but the model can also be interpreted as a model of motivated effort denial or career concerns (see Appendix B for more details). The information structure is straightforward in the case that the agent cares about the image that he conveys to the principal. The principal does not observe skill and observes effort with probability $1 - \nu$. In the case of self-image, the agent only imperfectly recalls how difficult the task was for him and how long he took for the task.[5] We thus assume that the agent knows his skill $\theta$ at date 1 but not anymore at date 2. To model effort denial, we let the agent at date 2 recall the time he spent on working on the task only with probability $1 - \nu$.[6] For example, the agent might not perfectly recall the time he started working, how long the lunch break was, or how many times he interrupted work for private matters.[7]

We summarize such image concerns as a linear function that is increasing in inferred skill.[8] Overall, an agent of type $\theta \in \{\theta_L, \theta_H\}$ has utility:

$$\begin{aligned} u(\theta, e, \hat{e}) &= W(y(\theta, e), \hat{e}) - c(e) - l_u \mathbb{1}_{e>\hat{e}} - l_o \mathbb{1}_{e<\hat{e}} \\ &\quad + s\left(\nu E[\theta | y(\theta, e), \hat{e}] + (1-\nu) E[\theta | y(\theta, e), e, \hat{e}]\right), \end{aligned} \tag{4}$$

where $s \geq 0$ is the weight on image concerns. An agent with $s = 0$ only cares for material payoff and not for his reputation. With probability $\nu$, beliefs about skill are based on the report $\hat{e}$; with probability $1 - \nu$ beliefs are in addition based on the actually exerted effort $e$. The parameters of the utility function $s, \nu, l_o$, and $l_u$ are also known to the principal.

[4] We derive the principal's optimal contract without these additional assumptions in the proof of Proposition 1. All results stay the same for the case that working overtime is efficient, but there is one parameter range where non-monotonic contracts are optimal because rewarding a failure can reduce overwork.

[5] For similar interpretations of imperfect recall, see e.g. Bénabou and Tirole (2004, 2006, 2011) and Battaglini et al. (2005).

[6] Letting the agent remember both skill and effort with probability $\nu$ yields identical results. In fact, one could equivalently assume that the agent learns his skill type only once he is working on the task and either finishes the task on time or not. Knowing skill is then directly tied to remembering effort. In Appendix B, we endogenize $\nu$ in terms of motivated effort denial.

[7] McNair (1991) and Sweeney and Pierce (2006) discuss the perceptions of how only "good hours" and not "bad hours" should be reported and how there is ambiguity in which hours are good and which are bad.

[8] Linear self-image utility rules out behavior that stems from information avoidance or information seeking analyzed for instance by Köszegi (2006) and Andolfatto et al. (2009). Moreover, the linear forms makes it possible to simply add image concerns that stem from self-esteem or from other sources like material consequences in the future.

The principal's profit is

$$\pi(y, \hat{e}) = y - W(y, \hat{e}). \tag{5}$$

With material outside payoffs equal to zero for both agent and principal, the agent values his outside option at $s\bar{\theta}$, which is consistent with either remaining ignorant about skill or expecting to learn it elsewhere.

Since the agent expects the same expected image gain regardless of whether he accepts or rejects the contract, the surplus that is generated in the work relationship is solely determined by the expected output, effort costs, and lying costs. We denote the maximum expected surplus by

$$S^* = py_H + (1-p)\max\{y_H - c, y_L\}. \tag{6}$$

To focus on the interesting cases in this model, we assume that image concerns are not too large. The maximum return from the task should be larger than the maximal image gain.

**Assumption 1.** *Assume throughout that* $l_u, l_o \leq c$ *and* $s\nu\Delta \leq y_H$.

The restriction on lying costs is only made for convenience since a lying cost of $c$ is already equivalent to perfect monitoring of effort. The second assumption ensures that the principal always offers a contract to the agent.

### 3.4 Solution Concept

The decisions that follow once the principal has offered a contract are all made by the agent but are interpreted as a game between the agent and his date-2-self or between the agent and the principal. The player who moves at date 2 only forms beliefs, about which the agent cares. A pure strategy of the agent assigns to each skill type an effort and a report. We denote such a strategy by $e(\theta) \in \{e_L, e_H\}$, $\hat{e}(\theta) \in \{e_L, e_H\}$. By $\mu(y, \hat{e})$ we denote the posterior belief (of agent or principal) that the agent's skill is $\theta_H$ when output is $y$ and report is $\hat{e}$.

For contracts that support multiple equilibria, we assume that the principal can choose an equilibrium. We assume, however, that the principal can only pick perfect Bayesian equilibria with reasonable belief systems. Specifically, we assume that beliefs have to fit the production technology and have to satisfy an extended D1 condition. By the former we mean that task failures generally have to be associated with low skills. This directly follows from the commonly known production technology, according to which any task failures must originate from low-skilled agents, no matter what the agent reports.[9]

By the latter we mean that we impose the D1 refinement (Cho and Kreps, 1987) in the setting with positive lying costs, and in the setting without lying costs only consider those equilibria that are the limits of D1 equilibria in the game with lying costs as lying costs go to zero.[10] Formally, an equilibrium $(e^*, \hat{e}^*, \mu^*)$ satisfies this refinement if the following is true:

[9] Beliefs $\mu(y_L, e_H), \mu(y_L, e_L) > 0$ are not structurally consistent (Kreps and Wilson, 1982) as there does not exist a behavior strategy that would lead to $y = y_L$ with positive probability and yield $\mu(y_L, e_H) > 0$ or $\mu(y_L, e_L) > 0$, respectively, as a consistent belief.

[10] Chen et al. (2008) use a similar idea to justify a refinement in the model of strategic information transmission of Crawford and Sobel (1982).

For lying costs $l = (l_o, l_u)$, let $\Gamma(l)$ be the signaling game that results once the agent has accepted a contract and the effort choice is fixed at $e = e^*$. This game hence consists only of nature drawing the agent's type, the effort report, and the belief updating of the date-2-player. The equilibrium $(e^*, \hat{e}^*, \mu^*)$ satisfies the refinement if there exists a sequence of lying costs $(l_n)_n$ with $l_n \neq (0,0)$ and $\lim_{n\to\infty} l_n = (0,0)$ and a corresponding sequence of signaling equilibria $(\hat{e}_n, \mu_n)_n$ of $\Gamma(l_n)$ that satisfy D1 with $\lim_{n\to\infty}(\hat{e}_n, \mu_n) = (\hat{e}^*, \mu^*)$.

The D1 refinement is quite strong: whenever some type would always benefit a little more from a deviation to an out-of-equilibrium signal, then the receiver, were he to observe this out-of-equilibrium-signal, must believe that it is infinitely more likely that it must have been sent by this type. In this model, it should mean that no matter how small the lying costs, an off-equilibrium-path belief must put probability one on the type that does not have to lie. While the refinement is strong, it merely pins down reporting behavior and does not matter for equilibrium payoffs.

# 4 Analysis

We first study the pure strategy equilibria of the signaling game that unfolds once the agent has accepted a contract $(w, b)$. The types of agent's behavior that we consider are *hidden, false, revealed*, and *no overtime*. In all these cases, the high type never works overtime, and the low-skilled agent works overtime in the first three cases. When overtime is hidden, the low type deflates his working hours in his report, whereas with false overtime the high type inflates his working hours. Under revealed overtime the agent reports truthfully. No overtime describes a behavior in which the low type does not work overtime and the types are separated by output.[11]

For these four cases of agent's potential behaviors, we specify the conditions on contracts that support them as equilibria that satisfy our refinement and make the agent accept the contract. Taking into account these incentive and participation constraints, we characterize the contracts that allow the principal to implement the different equilibria at lowest cost. We then turn to finding the principal's optimal contracts.

## 4.1 Agent's Behavior

### Hidden Overtime

By hidden overtime we refer to a behavior in which the high type works regular hours, the low type works overtime, and neither of the types reports overtime. That is, the agent's strategy in a hidden overtime equilibrium is $\hat{e}(\theta_L) = \hat{e}(\theta_H) = e_L$, $e(\theta_H) = e_L$ and $e(\theta_L) = e_H$. Both types complete the task and receive the bonus $b$. Moreover, the types pool such that following a task success and a regular-time report, both types obtain image equal to the expected skill $\bar{\theta}$ if effort remains unobservable. If effort is observed, also skills are inferred. Consequently,

[11] It is straightforward to see that we do not have to consider overtime work by the high type, for whom output is independent of effort: Working overtime would generate a cost of $c$ but yield at most a benefit of saving lying costs $l_o \leq c$. Moreover, the pure strategy profile in which the low-skilled agent completes the task but reports are reversed such that overtime reports are associated with high skills never arises given our assumptions (see Proposition 2 ). Proposition 2 also considers equilibria in mixed strategies.

on the equilibrium path, a high-skilled agent receives $b+s\nu\bar{\theta}+s(1-\nu)\theta_H$. A low-skilled agent in addition incurs overtime and lying costs and receives utility $b-c+s\nu\bar{\theta}+s(1-\nu)\theta_L-l_u$.

**Lemma 1.** *The hidden overtime equilibrium can be implemented only if* $s\nu p\Delta \geq l_u$. *The incentive constraints for the hidden overtime equilibrium are*

$$s\nu p\Delta \geq c-b+l_u \tag{7}$$

$$\textit{and } s\nu p\Delta \geq w+l_u. \tag{8}$$

*The participation constraint is*

$$b \geq (1-p)(c+l_u). \tag{9}$$

Condition (7) ensures that the low type works overtime. Working overtime together with hiding it allows the agent to generate a reputation benefit of $s\nu p\Delta$. This expected image gain needs to offset the net costs $c-b$ associated with working longer hours and the lying costs $l_u$ from hiding overtime. Condition (8) ensures that deviating to reporting overtime is not worthwhile because the image loss of an overtime report would outweigh the overtime pay and low type's lying costs. Note that overtime can be hidden even for positive overtime pay as long as image concerns are sufficiently important.

The degree of unobservability and imperfect recall of effort $\nu$ and uncertainty about skill $\Delta$ mediate the effects of image concerns. The more difficult it is for the principal to observe effort or for the agent himself to keep track of his working hours, the more room there is to deny effort, and thus the more likely overtime remains hidden. Similarly, there is more room to influence perceptions by hiding overtime when uncertainty about skills is large. Image concerns do not matter for the participation constraint (9). The agent only accepts the contract if the transfer, given equilibrium behavior, at least covers his expected effort and lying costs.

Let now $(b^{HO}, w^{HO})$ denote the contract that minimizes the cost of inducing hidden overtime. As no overtime compensation is paid, the overtime payment $w^{HO}$ can be zero or any $w \geq 0$ for which condition (8) holds. If $s\nu p\Delta < l_u$, the hidden overtime equilibrium cannot be implemented with any $w \geq 0$. The optimal bonus $b^{HO}$ is the lowest transfer that satisfies both the incentive constraint (7) and the participation constraint (9),

$$b^{HO} = \max\{(1-p)(c+l_u), c+l_u-s\nu p\Delta\}. \tag{10}$$

The optimal bonus to induce hidden overtime depends on the importance of image. Consider first the case that the maximal image gain is lower than the cost of overtime work, $s\nu\Delta \leq c+l_u$. In this case, the agent is not willing to work overtime without extrinsic incentives. Thus, the main challenge is to make the low-skilled agent exert effort. Because of the limited liability constraint this requires a larger compensation than making the agent sign the contract. If instead $s\nu\Delta > c+l_u$, the bonus is designed to ensure participation, and the principal receives the whole expected surplus.

**Revealed Overtime**

Next, consider the conditions for the high and low types to complete the task in regular time and overtime, respectively, and also to truthfully report their hours. Overtime is revealed if $e(\theta_L) = \hat{e}(\theta_L) = e_H$ and $e(\theta_H) = \hat{e}(\theta_H) = e_L$. In this case, the date-2-player (principal or agent) learns the underlying effort and thus also the type. The low type's payoff is $b+w-c+s\theta_L$ and the high type's payoff is $b+s\theta_H$. Beliefs following low output or an overtime report must put all probability on the low type, $\mu(y_L, e_H) = \mu(y_L, e_L) = \mu(y_H, e_H) = 0$.

**Lemma 2.** *The incentive constraints for the revealed overtime equilibrium are*

$$s\nu\Delta - l_u \leq w \leq s\nu\Delta + l_o, \tag{11}$$

$$\text{and } b + w \geq c. \tag{12}$$

*Participation requires*

$$b + (1-p)w \geq (1-p)c, \tag{13}$$

*which is implied by* (12).

Condition (11) is the truth-telling constraint. In particular, if there are no lying costs, then the overtime payment must equal the image loss that an overtime report entails. Condition (12) ensures that overtime payment and bonus together are high enough to cover the effort costs of the low-skilled worker.

To implement revealed overtime at minimal cost, the principal sets

$$b^{RO} = \max\{c - s\nu\Delta - l_o, 0\}. \tag{14}$$

If the limited liability constraint is binding such that $b^{RO} = 0$, then the least cost contract has

$$w^{RO} = \max\{c, s\nu\Delta - l_u\}, \tag{15}$$

else it has

$$w^{RO} = s\nu\Delta + l_o. \tag{16}$$

The total payment from principal to agent is given by

$$(1-p)w^{RO} + b^{RO} = \max\{(1-p)c, c - p(s\nu\Delta + l_o), (1-p)(s\nu\Delta - l_u)\}. \tag{17}$$

Consider the case that there are no lying costs. If the limited liability constraint $b \geq 0$ is not binding, which occurs if $s\nu\Delta \leq c$, the total expected transfer is the same as in the hidden overtime case. It is thus this limited liability constraint that destroys the equivalence between the two cases, making revealed overtime strictly worse for the principal than hidden overtime if $s\nu\Delta > c$.

**False Overtime**

False overtime captures the possibility that the high-skilled agent exaggerates working time. In this equilibrium, the low type works overtime and the high type regular hours, but both

report overtime. That is, $e(\theta_L) = \hat{e}(\theta_L) = e_H$, $e(\theta_H) = e_L$ and $\hat{e}(\theta_H) = e_H$. Again neither the output nor the self-report gives away the type of the agent, but this time it is the high type who might have to incur a cost by pretending to have worked overtime.

**Lemma 3.** *The incentive constraints for the false overtime equilibrium are*

$$w \geq s\nu(1-p)\Delta + l_o \tag{18}$$

$$\text{and } s\nu p\Delta \geq c - b - w. \tag{19}$$

*Participaton requires*

$$w + b \geq (1-p)c + pl_o, \tag{20}$$

*which is implied by (18) and (19).*

Reporting regular hours would create a more competent appearance and be a truthful report for the high type. However, reporting regular hours would also mean that the agent forgoes the overtime compensation. Condition (18) ensures that false overtime reporting is optimal. Condition (19) is the incentive constraint for the low-skilled agent, who only works long hours if the benefit of preserving his (self-)image exceeds the costs of doing so, which are given by the effort costs net of the transfer.

The least costly way to implement false overtime satisfies the condition

$$w^{FO} + b^{FO} = \max\{(1-p)c + pl_o, c - s\nu p\Delta, s\nu(1-p)\Delta + l_o\}. \tag{21}$$

Comparing equation (21) with the total transfer in the revealed overtime equilibrium, equation (17), one sees that false overtime is strictly more costly for the principal than revealed overtime if $l_o > 0$. If lying costs are zero, the total transfer to the agent is the same as in the revealed overtime case.

Apart from the lying costs, the difference to hidden overtime is due to the beliefs following an unexpected report. With hidden overtime, an unexpected overtime report triggers the belief that the agent is low-skilled, which is the optimal belief to sustain an equilibrium. With false overtime, an unexpected report of regular hours instead triggers the belief that the agent is high-skilled.

### No Overtime

No overtime describes a behavior of both types working regular hours, $e(\theta_L) = e(\theta_H) = e_L$. The high type completes the task and the low type does not. The agent's type can thus be inferred from the output. Reporting overtime might in this case simply be excluded by the principal or deterred by setting $w = 0$.[12] The low type has to prefer working regular hours to working overtime and pooling with the high type:

**Lemma 4.** *The no overtime equilibrium can be implemented if and only if* $s\nu\Delta \leq c + l_u$. *The*

[12] If reports are possible, off-equilibrium-path belief $\mu(y_H, e_H) = 0$ is consistent with our refinement and the overtime payment must satisfy $w \leq \min\{s\nu\Delta + l_o, c - b\}$.

*incentive constraint for the no overtime equilibrium is*

$$b + s\nu\Delta \leq c + l_u. \tag{22}$$

*The participation constraint is always met.*

If $s\nu\Delta \leq c + l_u$, the principal can prevent overtime work by setting

$$(b^{NO}, w^{NO}) = (0, 0), \tag{23}$$

which lets him capture the complete surplus $py_H + (1-p)y_L$. Else there is no contract for which the no overtime equilibrium exists.

## 4.2 Optimal Contracts

At date 0, the principal proposes a contract (and equilibrium behavior by the agent) that maximizes her payoff among all contracts that satisfy the limited liability and monotonicity constraints.

### No Recording System

As a benchmark, we first consider the case that there are no effort reports and the principal designs an optimal bonus $b^*$.

**Proposition 1.** *Assume that there are no effort reports. First suppose that working overtime is efficient, $y_H - y_L \geq c$. Both the surplus and principal's profit are weakly increasing in the strength of the image concerns $s\nu\Delta$. For $s\nu\Delta \geq c$, the principal's optimal contract induces overtime work, and the bonus is determined by a binding participation constraint, $b^* = (1-p)c$. For $s\nu\Delta < c$, there is no overtime work and $b^* = 0$ if $(1-p)(y_H - y_L - c) < p(c - s\nu\Delta)$, and overtime work and $b^* = c - s\nu p\Delta$ else. Only in this latter case, the agent receives strictly positive rent $p(c - s\nu\Delta)$ over his outside option payoff.*

*Now suppose that working overtime is inefficient, $y_H - y_L < c$. Both the surplus and the principal's profit are weakly decreasing in the strength of the image concerns $s\nu\Delta$. The agent always receives his outside option payoff. If $s\nu\Delta \leq c$, then there is no overtime work and $b^* = 0$. If $s\nu\Delta > c$, the principal's optimal contract induces overtime work with bonus $b^* = (1-p)c$.*

To assess the role of image concerns, start with the special case of an agent who does not care about image. If $s = 0$, making the agent work overtime means leaving a rent of $c$ to the high-skilled agent. If the gain from overtime work $y_H - y_L - c$ is positive but not large enough, the principal's optimal contract induces an under-provision of effort. Image concerns alleviate this inefficiency. As image concerns become more important, the rent that is needed to induce overtime work is reduced, and overtime becomes more and more attractive to the principal. Therefore, the surplus weakly increases as image concerns increase if overtime work is efficient.

In the parameter range $y_H - y_L \geq c$ it is structurally impossible for the agent to work too much from an efficiency point of view. In contrast, inefficient overwork can arise if

$y_H - y_L < c$.[13] When image concerns are high, maintaining image motivates the agent to work overtime, and it is impossible for the principal to contract the efficient level of effort. Although image concerns may make the low-skilled agent work too much, it is the principal who is hurt by this inefficiency: The agent rationally anticipates the potential overtime work and, therefore, will only accept a contract that compensates for the potential overtime work.

### Cheap Talk Reporting

We now suppose there is a system for reporting effort. Our first result concerns the case that reports are cheap talk. In standard problems with hidden effort and hidden effort costs, contracts condition only on output; additionally asking the agent for reports is not necessary. We show that this result carries over to our setting.

**Proposition 2.** *Assume that effort is reported and reports are cheap talk, $l_u = l_o = 0$. If $s\nu\Delta > c$, the principal strictly prefers the hidden overtime contract $(w^{HO}, b^{HO})$ over the revealed and false overtime contracts. In any optimal contract, overtime has to be hidden with positive probability. If $s\nu\Delta \leq c$ and $s\nu p\Delta < c - (y_H - y_L)(1-p)$, the no overtime contract $(w^{NO}, b^{NO})$ is optimal. Otherwise, the principal is indifferent between all contracts that induce the agent to work overtime. All payoffs are the same as in the case without reports (Proposition 1).*

Thus, the principal's optimal contract necessarily induces some hidden overtime when image concerns are important. For intuition, consider how the image concerns interact with the effort reports. Image concerns constitute a cheap way of providing incentives for effort provision. This is particularly true for the hidden overtime equilibrium, in which any deviation would result in a reputation for being low-skilled. In contrast, a deviation from a false overtime equilibrium would result in a reputation for being high-skilled. Moreover, when image concerns are relatively important, revelation of overtime requires a large overtime payment, while pooling the two types in the hidden overtime equilibrium allows the principal to reduce the agent's rent to zero despite the limited liability constraint. The optimal contract then sets overtime pay low enough to keep overtime at least partially hidden and induces participation with higher performance pay.

While Proposition 2 establishes that hidden overtime can occur if the agent is motivated by image, it is less clear whether this should be seen as problematic. The hidden overtime equilibrium does entail an inefficiency in the case $y_H - y_L < c$. However, it is not the case that the principal benefits from this inefficiency and, in that sense, exploits the agent. The agent's individual rationality bounds his payoff from below, and this bound is attained also when there is hidden overtime.

In the case $y_H - y_L \geq c$, there is no inefficiency in the parameter range where hidden overtime must be a consequence of the optimal contract. However, in this parameter range,

[13] While the standard agency problem is one of underprovision of effort, there are a few contributions that have also identified cases in which the optimal contract leads to overwork. Reasons include countervailing incentives (Lewis and Sappington, 1989), overemployment as a screening device (Sousa-Poza and Ziegler, 2003; Goldlücke and Schmitz, 2018), career concerns (Holmström, 1999; Andersson, 2002), or a salience bias that makes an agent focus too much on a potential bonus (Römeis et al., 2022).

the principal exploits the image concerns of the agent to reduce his rent to zero. The revealed or false overtime contracts would leave the agent with a rent instead. There is another difference between hidden and revealed overtime that may be interpreted as an exploitation of the agent from an ex-post perspective: In the hidden overtime equilibrium, the low-skilled agent does not receive full monetary compensation for his overtime work.

Proposition 2 also shows that revealing overtime with positive overtime pay can be part of an optimal contract if image concerns are not too important. Without image concerns, cheap talk reporting simply leads to overreporting. In contrast, with image concerns, overtime pay does not have to lead to false reporting since overtime reports carry an adverse signal about ability. This adverse signal discourages false overtime reporting, while the positive overtime pay may elicit a truthful overtime report even from an image concerned agent. While revealing overtime comes with the cost of overtime pay, truthful overtime reporting allows the principal to condition payments on overtime work, and this reduces the agent's information rents. However, Proposition 2 shows that revealed overtime is never strictly preferred to hidden overtime. If overreporting is costless, the revelation of overtime is restricted by the high type's incentives of overreporting, and so much so that the overtime pay cannot be set high enough for the better allocation of payments to dominate hiding overtime.

### Lying Costs

If the cost of an overtime report correlated with the hidden effort, reports would potentially be valuable to the principal as a signal of actually exerted effort. We now solve the model with general lying costs and find that once the agent incurs some costs when not telling the truth about hours worked, the principal can use the revealed overtime contract to reduce the agent's rent.

**Proposition 3.** *Assume that effort is reported and $l_o, l_u > 0$. If lying costs are large such that $c - l_o \leq s\nu\Delta \leq c + l_u$, the principal can implement the first best, with the revealed overtime contract if $y_H - y_L \geq c$ and the no overtime contract else. If image concerns are large such that $c + l_u < s\nu\Delta$, overtime is hidden with probability $\frac{p(s\nu\Delta - c - l_u)}{p(s\nu\Delta - c - l_u) + c + l_u}$. If image concerns are low such that $s\nu\Delta < c - l_o$, then the principal's preferred contract is the no overtime contract if $(1-p)(y_H - y_L - c) < p(c - s\nu\Delta - l_o)$, and else the revealed overtime contract.*

Figure 2 depicts the principal's profit for the pure strategy revealed, hidden, and no overtime contracts. When $s\nu\Delta \leq c - l_o$, lying costs do not alone prevent the high type's deviation to overreporting. If the principal induces overtime, the high type thus gets a rent, and the principal's profit increases in $s\nu\Delta$. When $c - l_o < s\nu\Delta \leq c + l_u$, the principal can set the performance-based component of the transfer to zero and compensate the cost of overtime work for the low type by setting $w = c$. Overtime pay can thus be set to reveal the hidden action, and the contract can be conditioned on effort. When $c + l_u < s\nu\Delta$, revealing overtime from the low type becomes exceedingly expensive. As the low type's image concerns are strong enough to alone incentivize effort, the moral hazard problem disappears and, thus, the only benefit of revealing overtime is saving the lying costs. The optimal contract minimizes the lying costs subject to the constraint that the agent receives no rent.

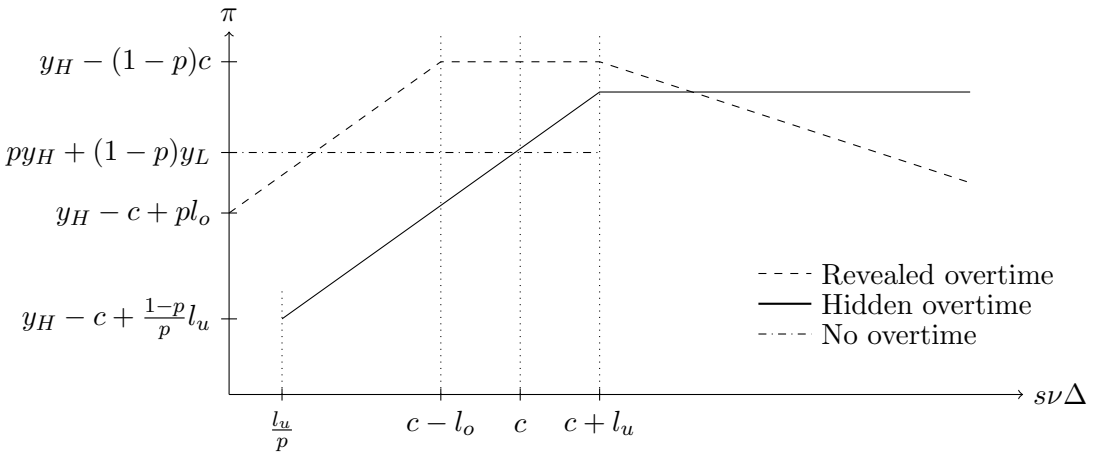


Figure 2: Principal's profit for $l_u, l_o > 0$, $y_H - y_L > c$.

Hence, asking for reports and inducing revealed overtime is strictly optimal for the principal when image concerns are not too high. Stricter constraints on overreporting hours in the form of lying costs increase the principal's leeway in targeting the compensation toward incentivizing overtime work of the low type and away from the rents of the high type. Nevertheless, when image concerns are high, the principal prefers to induce some hidden overtime.

### 4.3 Monitoring

With Proposition 3, we can study the effects of implementing a recording system. We assume that any recording system is characterized by its lying costs, which measure how easy it is to manipulate the system. The lying costs can also have an agent-specific component. For example, the effect of making the agent hand in signed report cards depends on the agent's aversion toward lying and signing inaccurate statements. We compare the effect of installing a system with specific lying costs to the benchmark of having no recording system at all, which is equivalent to a recording system that induces zero lying costs. Hence, installing a recording system with positive lying costs can also be interpreted as there initially being a cheap talk recording system and implementing monitoring to increase lying costs.

Consider first a perfect monitoring system. If the agent always reported honestly, a cheap talk recording system would already have this property.

**Corollary 1.** *With perfect monitoring, which is achieved by lying costs $l_u \geq s\nu\Delta - c$ and $l_o \geq c - s\nu\Delta$, the principal's first best is obtained with contract $(b^{RO}, w^{RO}) = (0, c)$. Installing such a recording system weakly increases the surplus and principal's profit and weakly decreases the agent's expected payoff.*

It seems likely that every recording system must remain imperfect and perfect monitoring is rarely attainable. Consider next the case where every monitoring system can be manipulated by reporting more hours than actually worked. Pretending to have worked less than the actually worked hours is often easily deterred. For example, if the agent has to be present at the facility in order to do his work, even a simple recording system can make it difficult for him to falsely claim that he was not there, while he might still find ways to spend his time

at the facility with other activities. Similarly, if the agent is only able to work if he is logged on to some technical system, he can easily report higher but not lower working time than the actual number of hours.

**Corollary 2.** *Assume that understating hours worked is too costly, $l_u > s\nu\Delta - c$, but over-reporting hours worked is easy, $l_o = 0$. Installing the recording system weakly increases the surplus and principal's profit and leaves the agent's expected payoff unchanged.*

This follows directly from Proposition 3: If $s\nu\Delta \leq c$, the principal chooses between revealed overtime and no overtime, and payoffs are the same as in the cheap talk case. If $s\nu\Delta > c$, a revealed overtime equilibrium with contract $(b^{HO}, w^{HO}) = (0, c)$ obtains if $y_H - y_L \geq c$, otherwise the no overtime contract $(b^{NO}, w^{NO})$ results. A recording system that completely deters underreporting thus weakly increases efficiency. Inefficient overtime work does not occur anymore. If overtime work is efficient, the principal gets the whole expected surplus as in the cheap talk case, but the low-skilled agent is compensated for the cost of his overtime work.

What if the principal contracts with an agent who will feel bad about falsely reporting overtime and collecting an overtime payment? We now consider the case $0 = l_u < l_o$. Social conventions may judge underreporting as less reprehensible than overreporting.[14] A technical system may also exhibit these asymmetric lying costs: If the agent can log on to a system that monitors his activities (for example, by tracking mouse movements and keystrokes), but he is also able to work offline and transfer the finished work to the system, then it is more difficult to report higher than lower working time.

**Corollary 3.** *Suppose $0 = l_u < l_o < c$. Installing the recording system weakly increases the surplus and principal's profit.*

Again, this follows directly from Proposition 3: For $s\nu\Delta < c - l_o$, the principal reveals overtime if $sp\Delta > c - (1-p)(y_H - y_L) - pl_o$ and otherwise the no overtime contract obtains. For $c - l_o < s\nu\Delta < c$, the principal reveals overtime if $y_H - y_L > c$ and otherwise the no overtime contract obtains. For $s\nu\Delta > c$, the lying cost $l_o$ has no effect.

Hence, a cost of exaggerating hours worked only affects the moral hazard problem and not the hidden overtime problem. An agent who values honesty might be worse off if working time is recorded because the rent that the agent receives is decreasing in the lying costs $l_o$. For such an agent, the recording system would essentially be a monitoring system that allows the principal to make part of the wage contingent on effort. It is, however, also possible that the agent benefits from this effect. This happens if in the absence of the recording system, there would be underprovision of effort with the principal implementing no overtime.

# 5 Regulation

In 2019, the European Court of Justice established an obligation for employers to record the daily working times of their employees. Member states have to implement this obligation

[14] Barrainkua and Espinosa-Pike (2015) find that underreporting of working hours is often considered neither unethical nor ethical. Overreporting or "time theft", on the other hand, is usually considered a crime (Henle et al., 2010; Harold et al., 2022; Hu et al., 2023).

and fill in the details on the required characteristics of such a recording system. We will use our model to compare the situation with a recording system to the one without and evaluate whether regulation can eliminate inefficiencies. Of course, the policymaker's goal might not only be efficiency but also the protection of the employee who is seen as the weaker party. We therefore also consider other objectives.

Our model with zero lying costs corresponds to a situation in which a recording system is easily manipulated. This may be the case, for example, if employees work from home. As was shown in Proposition 2, in this case, an obligation to report hours worked has no effect since the agent may (self-)deceive with his reports. An obligation to install such a recording system leaves principal's and agent's payoff the same, except for any costs associated with the system. One might think that a recording system has an effect if the court also imposes a minimum overtime pay. However, hidden overtime is consistent with some compensation for overtime work as long as (8) is satisfied.

While it might not be very surprising that simply requiring reports does not have an effect, the European Court of Justice has specified some minimal requirements on the recording system in its 2019 ruling. Working time is supposed to be recorded with an "objective, reliable and accessible system" (ECJ, C55/18, 2019, paragraph 60). We therefore ask what additional requirements on the recording system, for example, on the susceptibility to manipulation measured by the lying costs, could lead to an improvement.

According to Proposition 3, installing a recording system with positive lying costs has ambiguous effects on the agent's and principal's payoff. It can lead to a switch from hidden overtime to no overtime, thereby solving the problem of hidden overwork. It can benefit both principal and agent, but only if it leads to a switch toward overtime. It can reduce the agent's rent with increased surveillance. It can also be merely a bureaucratic measure with no effect other than increased transaction costs.

Note that even if a monitoring system increases welfare, this does not mean that its use has to be mandatory. For regulation to be necessary, it has to be the case that there is an inefficiency that can only be addressed by a third party. Hence, we do not only have to ask whether an improvement relative to the situation without reports is possible, but also whether the contracting parties themselves cannot achieve this increase in efficiency.

To assess the overall efficiency of requiring effort reports, we take into account the possible costs of using such a system. Apart from the agent's lying costs, there may be a set-up cost $k_P \geq 0$ borne by the principal and a reporting cost $k_A \geq 0$ borne by the agent.

**Proposition 4.** *Assume that a recording system with lying costs $l_u \geq 0$ and $l_o = 0$ and reporting costs $k_A \geq 0$ is available at set-up cost $k_P \geq 0$. The principal installs the system if and only if installing it leads to an increase in efficiency.*

Thus, the principal has sufficient incentives to install recording systems that are designed to reduce hidden overtime. If the principal does not install such a monitoring system, then the reason is that the associated costs outweigh the efficiency gain. An external obligation to install such a monitoring system would either be not necessary or inefficient.

Since Proposition 4 deals only with reporting systems with $l_o = 0$, recording systems with $l_o > 0$ need a closer look to answer the question of optimal regulation. Those are systems that

aim at increased surveillance, either through the use of electronic devices or monitoring by the principal. The proof of Proposition 4 shows that the principal sometimes fails to install such a recording system although it could efficiently reduce moral hazard. In that case, requiring the principal to adopt a recording system can lead to an increase in efficiency. It may, however, also be in the principal's interest to implement a system with $l_o > 0$ merely to capture more rent from the agent. The principal can thus have excessive as well as insufficient incentives to install such a system. Therefore, we now explore the consequences of giving the agent some right to require a recording system.

If the agent has a right to co-determine the design of a recording system, it must be a right that cannot be forfeited with the signing of the contract. At the same time, the principal should be allowed to make adjustments to overtime payment when there is a new system. Consequently, we assume the following timing: An agent who has accepted the contract can demand a recording system, and the principal can react to this decision by offering a new contract that the agent can accept or reject. This means that if the agent uses his right to change the recording system, the old contract is void. The agent's decision has to be made before he knows how difficult he will find the task.

**Proposition 5.** *Assume the principal is not obligated to install a recording system and that the agent has the right to reject any system with $l_o > 0$, but he can also require a system with $l_o > 0$ to be installed (even if there is a system with lower lying costs already in place). With this rule, a recording system is installed if and only if it is efficient to do so.*

The agent's power to co-determine the recording system effectively leads to a bargaining solution ensuring that no system decreasing either agent's or principal's payoff is installed. When the agent, with the right to determine the surveillance level of the recording system, participates in the decision to install a recording system, the system will be adopted if and only if it increases efficiency. Compared to this optimal regulatory regime, a strict obligation to install a recording system does not achieve an outcome adapted to different circumstances.

Since an obligation to install a recording system seldom increases efficiency, we briefly consider other reasons for a policymaker to wish for such a regulation. In its ruling, the European Court of Justice states that "the effective protection of the safety and health of workers should not be subordinated to purely economic consideration" (ECJ, C55/18, 2019, paragraph 66). We will therefore consider a policymaker who does not care about the cost of the recording system and aims at eliminating inefficient overtime and ensuring adequate monetary compensation for efficient overtime work. As shown in Corollary 2, there is one regulation that achieves this without any negative consequences (other than the cost of keeping the records). Requiring a recording system that successfully deters hiding overtime increases efficiency by eliminating overwork without hurting the agent.

While this recording system leads to a Pareto improvement (if the administrative costs can be neglected), the agent's payoff is not increased. A positive rent for the agent might be achieved if the principal faced adverse consequences for hidden overtime. For example, there could, with some probability, be an investigation by a higher authority leading to a fine if overtime had remained hidden. A threat to monitor the truthfulness of reports will make revealing overtime more attractive and, therefore, weakly increase the agent's rent.

However, if the total administrative costs of a mandatory recording system become too large, an obligation to install it may also lead to an inefficient shutdown.

In practice, monitoring by a labor inspectorate is likely to lead to additional transaction costs. The monitoring problems should indeed be larger for an external authority or a worker's council than for the employer. Moreover, if the principal is punished for false reports, she might be tempted to install a more capable monitoring system. A surveillance system that achieves perfect monitoring leaves the principal with the whole surplus and also generates the required records. The obligation to record the agent's working time would hence be beneficial to the agent only if records can be kept in a way that they can be inspected by some third party but at the same time, the recording system can still be manipulated by the agent.

# 6 Concluding Remarks

We have explained hidden overtime in terms of signaling ability, career concerns, self-deception, and effort denial. Requiring extra time to complete a task signals low ability. An overtime report may thus be interpreted by the principal or the agent himself as inability to complete the task during regular working hours. This, on the one hand, disciplines the agent's overtime reports such that honest reporting may arise in equilibrium but, on the other hand, motivates the agent to deny having worked overtime. We have shown how overtime pay contingent on self-reported hours can be part of an optimal contract when the agent has image concerns but how overtime remains hidden when image concerns are high.

Hiding overtime allows low-skilled agents to pool with high-skilled agents leading to a positive correlation between underreporting hours and self-esteem and performance evaluations as documented by Akers and Eaton (2003) and Coulshed et al. (2022). The role of the unobservability of effort in facilitating hiding overtime may be behind the joint increase in teleworking and hidden overtime (ILO, 2017). The model also predicts imperfect recall of effort to increase the prevalence of hidden overtime. Therefore, the time lag between effort and its report may increase the propensity of hiding overtime work, as found in the experiment of Reffett et al. (2014). Similarly, the lack of boundaries between leisure and work time, leading to ambiguity about the exact hours worked, may increase the prevalence of hidden overtime. The model predicts that uncertainty about skills makes hiding overtime more likely and thus may explain why younger employees more often hide overtime than older employees, as observed by Shapeero et al. (2003).

The inefficiency generated by hidden overwork might call for policy interventions. We have shown that a requirement to record working time can mitigate the hidden overtime problem if understating hours worked is made difficult, and the moral hazard problem if overstating hours worked is made difficult. It is also possible that the only consequence of a recording system are the set-up and recording cost associated with such a system. This is likely to be the case in occupations in which (self-)image concerns are pronounced. Hence, there are cases in which the benefits of a recording system are larger than the associated costs and cases in which they are lower. To achieve an efficient outcome in all these cases, regulation has to be either very detailed or simply non-existent: The principal would voluntarily install a system that can reduce the hidden overtime problem if the cost of the system is lower than the

increase in the surplus. A legal obligation to install such a system is not necessary. Even in cases where such an obligation would be beneficial, which can occur for systems that reduce monitoring problems, it suffices if the agent has the right to require or reject a recording system.

The reason behind this result is that the agent is rational and is therefore compensated for expected overtime work, at least from an ex ante perspective. The policymaker could have a more naïve employee in mind who does not foresee that image concerns will make him work more if the task turns out to be difficult. Similarly, an overconfident agent may not consider the possibility that he might find the task difficult and, therefore, in signing the contract, require no compensation for the additional work that might be needed. A recording system that prevents reports that are lower than actual working time ensures adequate compensation for costly overtime work preventing exploitation of a naïve employee.

Another factor that may play a role in the obligation to record working time is control aversion. Employees may dislike the loss of autonomy that constant recording of working time entails, which may be an argument against any such record-keeping system. However, this hidden cost of control (Falk and Kosfeld, 2006; Akerlof and Kranton, 2008) is partly based on reciprocity (Burdin et al., 2018). A legal obligation to control may help the principal since if employees know the decision was beyond their employer's authority, they might not feel the need to reciprocate negatively.

There are other potentially relevant factors that our model does not address. One is that if output is not perfectly observable, the same image concerns that lead to overtime work in our model might lead instead to reporting an unfinished project as completed. Another is the possibility that has been highlighted in previous research, that high effort might also be a signal of low marginal cost of effort. Studying the determinants and interplay of the two effects is an interesting question for future research. Another avenue for future research concerns the dynamic aspects of the labor relationship. Although image concerns capture some dynamic effects in reduced form, others are not produced by our model, for instance, dynamic effects with respect to opportunities for task redesign or task choice. If agents always honestly reported overtime, the principal would, over time, learn to assign the most profitable tasks to the agents. Gaining information about production costs would be an additional incentive for the principal to reveal overtime work.

We have interpreted the agent's skill or task difficulty resulting from a match between the agent and the task. An alternative is to have separate parameters for the agent's skill and task difficulty and uncertainty of whether perceived difficulty is due to the task or due to the agent. This could be especially interesting in a model with multiple agents where task difficulty is common to all agents, but the agents' skills vary. Strong peer effects may arise when not being able to complete a task given others' completion of the same task is a strong signal of low skills.

## References


Akerlof, G. (1976). The economics of caste and of the rat race and other woeful tales. *The Quarterly Journal of Economics*, 90(4):599–617.

Akerlof, G. A. and Kranton, R. E. (2008). Identity, supervision, and work groups. *American Economic Review*, 98(2):212–17.

Akers, M. D. and Eaton, T. V. (2003). Underreporting of chargeable time: The impact of gender and characteristics of underreporters. *Journal of Managerial Issues*, 15(1):82–96.

Akers, M. D., Eaton, T. V., and Horngren, C. T. (1999). Underreporting chargeable time: A continuing problem for public accounting firms. *Journal of Applied Business Research*, 15(1):13–20.

Andersson, F. (2002). Career concerns, contracts, and effort distortions. *Journal of Labor Economics*, 20(1):42–58.

Andolfatto, D., Mongrain, S., and Myers, G. (2009). Rational truth-avoidance and self-esteem. *Canadian Journal of Economics/Revue canadienne d'économique*, 42(1):141–154.

Anger, S. (2008). Overtime work as a signaling device. *Scottish Journal of Political Economy*, 55(2):167–189.

Aron, D. J. (1987). Worker reputation and productivity incentives. *Journal of Labor Economics*, 5(4, Part 2):S87–S106.

Barrainkua, I. and Espinosa-Pike, M. (2015). New insights into underreporting of time: the audit partner context. *Accounting, Auditing and Accountability Journal*, 28(4):494–514.

Battaglini, M., Bénabou, R., and Tirole, J. (2005). Self-control in peer groups. *Journal of Economic Theory*, 123(2):105–134.

Bénabou, R. (2015). The economics of motivated beliefs. *Revue d'économie politique*, 125(5):665–685.

Bénabou, R. and Tirole, J. (2002). Self-confidence and personal motivation. *The Quarterly Journal of Economics*, 117(3):871–915.

Bénabou, R. and Tirole, J. (2004). Willpower and personal rules. *Journal of Political Economy*, 112(4):848–886.

Bénabou, R. and Tirole, J. (2006). Incentives and prosocial behavior. *American Economic Review*, 96(5):1652–1678.

Bénabou, R. and Tirole, J. (2009). Over my dead body: Bargaining and the price of dignity. *American Economic Review*, 99(2):459–65.

Bénabou, R. and Tirole, J. (2011). Identity, morals, and taboos: Beliefs as assets. *The Quarterly Journal of Economics*, 126(2):805–855.

Bernheim, B. D. and Thomadsen, R. (2005). Memory and anticipation. *The Economic Journal*, 115(503):271–304.

Burdin, G., Halliday, S., and Landini, F. (2018). The hidden benefits of abstaining from control. *Journal of Economic Behavior & Organization*, 147:1–12.

Chen, Y., Kartik, N., and Sobel, J. (2008). Selecting cheap-talk equilibria. *Econometrica*, 76(1):117–136.

Cho, I.-K. and Kreps, D. M. (1987). Signaling games and stable equilibria. *The Quarterly Journal of Economics*, 102(2):179–221.

Coulshed, A., Fernandes, B., and Hettige, S. (2022). Overtime claiming among australian doctors-in-training. *Australian Health Review*, 46(2):163–169.

Crawford, V. P. and Sobel, J. (1982). Strategic information transmission. *Econometrica*, 50(6):1431–1451.

ECJ, C55/18 (2019). Case c-55/18 CCOO v Deutsche Bank. Judgement of the Court (Grand Chamber) of 14 May 2019. European Court of Justice.

Falk, A. and Kosfeld, M. (2006). The hidden costs of control. *American Economic Review*, 96(5):1611–1630.

Fernbach, P. M., Hagmayer, Y., and Sloman, S. A. (2014). Effort denial in self-deception. *Organizational Behavior and Human Decision Processes*, 123(1):1–8.

Gibbons, R. and Murphy, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy*, 100(3):468–505.

Goldlücke, S. and Schmitz, P. W. (2018). Pollution claim settlements reconsidered: Hidden information and bounded payments. *European Economic Review*, 110:211–222.

Harold, C. M., Hu, B., and Koopman, J. (2022). Employee time theft: Conceptualization, measure development, and validation. *Personnel Psychology*, 75(2):347–382.

Henle, C. A., Reeve, C. L., and Pitts, V. E. (2010). Stealing time at work: Attitudes, social pressure, and perceived control as predictors of time theft. *Journal of Business Ethics*, 94:53–67.

Holmström, B. (1999). Managerial incentive problems: A dynamic perspective. *The Review of Economic Studies*, 66(1):169–182.

Hu, B., Harold, C., and Kim, D. (2023). Stealing time on the company's dime: Examining the indirect effect of laissez-faire leadership on employee time theft. *Journal of Business Ethics*, 183:475—-493.

Hvide, H. K. (2002). Pragmatic beliefs and overconfidence. *Journal of Economic Behavior & Organization*, 48(1):15–28.

ILO (2017). *Working anytime, anywhere: The effects on the world of work*. International Labour Organization.

Jones, P., Bano, N., and Cole, M. (2021). *The Right to Disconnect*. Autonomy.

Köszegi, B. (2006). Ego utility, overconfidence, and task choice. *Journal of the European Economic Association*, 4(4):673–707.

Kőszegi, B. and Li, W. (2008). Drive and talent. *Journal of the European Economic Association*, 6(1):210–236.

Kreps, D. M. and Wilson, R. (1982). Sequential equilibria. *Econometrica*, 50(4):863–894.

Landers, R. M., Rebitzer, J. B., and Taylor, L. J. (1996). Rat race redux: Adverse selection in the determination of work hours in law firms. *The American Economic Review*, 86(3):329–348.

Lewis, T. R. and Sappington, D. E. (1989). Countervailing incentives in agency problems. *Journal of Economic Theory*, 49(2):294–313.

Löwenstein, G. (1987). Anticipation and the valuation of delayed consumption. *The Economic Journal*, 97(387):666–684.

McNair, C. (1991). Proper compromises: The management control dilemma in public accounting and its impact on auditor behavior. *Accounting, Organizations and Society*, 16(7):635–653.

Mijović-Prelec, D. and Prelec, D. (2010). Self-deception as self-signalling: a model and experimental evidence. *Philosophical Transactions of the Royal Society B*, 365:227–240.

Otley, D. T. and Pierce, B. J. (1996). Auditor time budget pressure: consequences and antecedents. *Accounting, Auditing & Accountability Journal*, 9(1):31–58.

Quattrone, G. A. and Tversky, A. (1986). Self-deception and the voter's illusion. In Elster, J., editor, *The Multiple Self*, pages 35–38. Cambridge University Press, Cambridge.

Reffett, A., Eaton, T. V., and Gannod, G. (2014). An experimental examination of underreporting of time. *Journal of Managerial Issues*, 26(2):112–129.

Römeis, F., Herweg, F., and Müller, D. (2022). Salience bias and overwork. *Games*, 13(1):15.

Sampson, A. A. (2002). Weekenders and workaholics. *European Journal of Political Economy*, 18(1):193–208.

Schelling, T. C. (1986). The mind as a consuming organ. In Elster, J., editor, *The Multiple Self*, pages 177–195. Cambridge University Press, Cambridge.

Schwardmann, P. and van der Weele, J. (2019). Deception and self-deception. *Nature Human Behaviour*, 3(10):1055–1061.

Shapeero, M., Koh, H. C., and Killough, L. N. (2003). Underreporting and premature sign-off in public accounting. *Managerial Auditing Journal*, 18(6/7):478–489.

Sloman, S. A., Fernbach, P. M., and Hagmayer, Y. (2010). Self-deception requires vagueness. *Cognition*, 115(2):268–281.

Smith, W. R., Hutton, M. R., et al. (1995). Underreporting time: An analysis of current tax practice. *Journal of Applied Business Research*, 11(1):39–45.

Sousa-Poza, A. and Ziegler, A. (2003). Asymmetric information about workers' productivity as a cause for inefficient long working hours. *Labour Economics*, 10(6):727–747.

Spence, M. (1973). Job market signaling. *The Quarterly Journal of Economics*, 87(3):355–374.

Sweeney, B. and Pierce, B. (2006). Good hours, bad hours and auditors' defence mechanisms in audit firms. *Accounting, Auditing & Accountability Journal*, 19(6):858–892.

Trivers, R. (2011). *Deceit and self-deception: Fooling yourself the better to fool others*. Penguin UK.

van der Weele, J. J. and von Siemens, F. A. (2020). Bracelets of pride and guilt? An experimental test of self-signaling. *Journal of Economic Behavior & Organization*, 172:280–291.

# A Proofs

*Proof of Lemma 1.* We first show that with hidden overtime, D1 restricts $\mu(y_H, e_H) = 0$. Intuitively, if the low type deviates to revealing his overtime, he stops lying whereas if the high type deviates to overreporting he starts lying. Thus, the deviation is more costly for the high type. Fix $e^*(\theta_L) = e_H$ and $e^*(\theta_H) = e_L$ and let $u^*(\theta)$ denote the expected equilibrium payoff of a type $\theta$ sender. Then define

$$D(\hat{e}, \theta) = \{\mu : u(\theta, e^*(\theta), \hat{e}, \mu) > u^*(\theta)\}$$

as the set of inferences that lead to payoffs that are greater than the equilibrium payoff for type $\theta$, given a report $\hat{e}$.[15] A reporting strategy $\hat{e}^*$ together with beliefs $\mu^*$ satisfies D1 if for any unsent message $\hat{e}$ it holds that if $D(\hat{e}, \theta_H) \subset D(\hat{e}, \theta_L)$ (proper subset), then $\mu^*(y_H, \hat{e}) = 0$ and if $D(\hat{e}, \theta_L) \subset D(\hat{e}, \theta_H)$ (proper subset), then $\mu^*(y_H, \hat{e}) = 1$.

Suppose $0 < l_o$ or $0 < l_u$. Note that $\mu \in D(e_H, \theta_L)$ means

$$s\nu(\mu\theta_H + (1-\mu)\theta_L) + w > s\nu\bar{\theta} - l_u. \tag{24}$$

Since for example $s\nu(1-p)\Delta + w > -l_u$, it holds that $D(e_H, \theta_L) \neq \emptyset$. Note also that $\mu \in D(e_H, \theta_H)$ means

$$s\nu(\mu\theta_H + (1-\mu)\theta_L) + w - l_o > s\nu\bar{\theta}.$$

This condition is more difficult to satisfy than (24), and therefore $D(e_H, \theta_L) \subset D(e_H, \theta_H)$ (proper subset). Then by D1 we have $\mu(\theta_H, e_H) = 0$.

Hence, beliefs following a deviation to an overtime report or low output put probability one on the low type, $\mu(y_L, e_H) = \mu(y_L, e_L) = \mu(y_H, e_H) = 0$. With these beliefs, a low type agent who deviates to low effort receives $s\theta_L$. Rearranging the condition $b - c + s\nu\bar{\theta} + s(1-\nu)\theta_L - l_u \geq s\theta_L$ yields (7). Lying to hide overtime is weakly more difficult to incentivize than reporting regular hours truthfully. If the low-skilled agent reported overtime instead, he would receive $b + w - c + s\theta_L$, which is lower than his hidden overtime payoff if and only if condition (8) is met.

The equilibrium for zero lying costs satisfies our extended D1 refinement as it is a limit of D1 equilibria for positive lying costs. Note also that a non-monotonic contract would not help here in implementing this behavior of the agent in a cheaper way. □

*Proof of Lemma 2.* In text. □

*Proof of Lemma 3.* Suppose $0 < l_o$ or $0 < l_u$. We first show that with false overtime, D1 restricts $\mu(y_H, e_L) = 1$. Intuitively, a deviation to reporting regular hours is less costly for the high type who stops lying than for the low type who starts lying. Let $u^*(\theta)$ denote the

[15] Note that because our model has an interpretation of the agent signaling to the principal, we can apply D1 to a game in which the principal responds to a belief in a way that creates the image utility for the agent. As a shortcut we equate the principal's best response action to a belief with the belief and make the dependence of the agent's utility on the belief explicit in the notation.

expected equilibrium payoff of a type $\theta$ sender and

$$D(\hat{e}, \theta) = \{\mu : u(\theta, e^*(\theta), \hat{e}, \mu) > u^*(\theta)\}.$$

Type $\theta_L$ is deleted for strategy $\hat{e} = e_L$ under criterion D1 if $D(e_L, \theta_L) \subset D(e_L, \theta_H)$, where this inclusion has to hold strictly. Note that $\mu \in D(e_L, \theta_H)$ means

$$s\nu(\mu\theta_H + (1-\mu)\theta_L) > s\nu\bar{\theta} + w - l_o \tag{25}$$

Hence, $D(e_L, \theta_L) \neq \emptyset$ is equivalent to $s\nu(1-p)\Delta > w - l_o$. Note also that $\mu \in D(e_L, \theta_L)$ means

$$s\nu(\mu\theta_H + (1-\mu)\theta_L) - l_u > s\nu\bar{\theta} + w$$

Note that this condition is more difficult to satisfy than (25). Therefore, under the condition $s\nu(1-p)\Delta > w - l_o$, $D(e_L, \theta_H) \subset D(e_L, \theta_L)$ (proper subset). Then by D1 we have $\mu(\theta_H, e_L) = 1$. If $s\nu(1-p)\Delta \leq w - l_o$, then both types choose to report overtime no matter what beliefs are.

If the agent follows the false overtime strategy, the low type's utility is $b + w - c + s\nu\bar{\theta} + s(1-\nu)\theta_L$ and the high type's utility is $b + w - l_o + s\nu\bar{\theta} + s(1-\nu)\theta_H$. Given off-equilibrium-path beliefs $\mu(y_H, e_L) = 1$ and $\mu(y_L, e_H) = \mu(y_L, e_L) = 0$, the high-type-agent reports overtime if and only if $w + s\nu\bar{\theta} - l_o \geq s\nu\theta_H$, which is condition (18). The low type works overtime if and only if $b + w - c + s\nu\bar{\theta} + s(1-\nu)\theta_L \geq s\theta_L$, which is condition (19).

The equilibrium for zero lying costs satisfies our extended D1 refinement as it is a limit of D1 equilibria for positive lying costs for underreporting, $l_u \to 0$ and $l_o = 0$.

To see that (18) and (19) imply the participation constraint, note that those two conditions imply that $w + b \geq s\nu(1-p)\Delta + l_o$ and $b + w \geq c - s\nu\Delta p$, which implies that

$$w + b \geq p((1-p)\Delta + l_o) + (1-p)(c - s\nu\Delta p) = (1-p)c + pl_o.$$

□

*Proof of Lemma 4.* In text. □

*Proof of Proposition 1.* We consider here also a more general class of contracts without the monotonicity restrictions to show how these simplifications affect the results.

Step 1 (Optimal bonus contract)
Let $W(y_H)$ and $W(y_L)$ be the payments given high and low output, respectively. The limited liability constraint on payments is $W(y_H), W(y_L) \geq 0$. The cheapest contract that induces both types to pool on high output and accept the contract is $W(y_L) = 0$ and $W(y_H) = \max\{c - s\nu p\Delta, (1-p)c\}$. This is essentially the hidden overtime contract for zero lying costs, with payoffs $y_H - \max\{c - s\nu p\Delta, (1-p)c\}$ for the principal and $s\bar{\theta} + \max\{0, p(c - s\nu\Delta)\}$ for the agent. If $s\nu\Delta > c$ and $y_H - y_L \geq c$, there is clearly no better contract, since the principal receives the full surplus $S^*$.

The cheapest contract that induces $e(\theta_L) = e_L$ is the no overtime contract $W(y_H) = W(y_L) = 0$. However, the principal can make the low type work regular hours only if $s\nu\Delta \leq c$.

If $s\nu\Delta \leq c$ and $y_H - y_L \leq c$, there is clearly no better contract as the principal again receives $S^*$.

Consider now the case $s\nu\Delta \leq c$ and $y_H - y_L > c$. Inducing a mixed strategy is dominated by either full separation (no overtime) or pooling (hidden overtime), so that we have to compare the principal's profit in these cases. The pooling contract is optimal if $y_H - (c - s\nu p\Delta) \geq py_H + (1-p)y_L$. Rearranging yields $s\nu p\Delta \geq c - (y_H - y_L)(1-p)$.

Finally, consider the case $s\nu\Delta > c$ and $y_H - y_L < c$. While inducing a mixed strategy is typically dominated by one of the corner cases of the low type working never or always overtime, if no overtime is impossible, mixing between the two effort levels might be an outcome that the principal wants to implement. With the restriction $0 = W(y_L) \leq W(y_H)$, however, the low type will necessarily work overtime and the optimal contract must be the hidden overtime contract. In this parameter range, the outcome is inefficient because the agent works too much.

Consider now the case without this restriction on contracts and assume that also non-monotonic contracts are possible. Let $\epsilon \in [0,1]$ be the probability with which the low type chooses high effort, and let $\tilde{p} = p + (1-p)\epsilon$. The low type is indeed indifferent between high and low effort if $W(y_L) = W(y_H) - c + s\nu\frac{p}{\tilde{p}}\Delta$. With limited liability, a minimal $\tilde{p}$ is obtained for $W(y_H) = 0$. The agent has to be compensated for the increase in working time so that the participation constraint binds: $(1-\tilde{p})W(y_L) = (\tilde{p} - p)c$. This means that $\tilde{p}$ can be found by solving

$$(1-\tilde{p})(-c + s\nu\frac{p}{\tilde{p}}\Delta) = (\tilde{p} - p)c,$$

which yields

$$\tilde{p} = \frac{s\nu p\Delta}{s\nu p\Delta + (1-p)c} \text{ and } \epsilon = \frac{p(s\nu\Delta - c)}{c + p(s\nu\Delta - c)}.$$

The principal receives $y_L + \tilde{p}(y_H - y_L) - (\tilde{p} - p)c$. In this mixing contract, inefficient overtime work occurs with some probability (and if it occurs, it is hidden).

Step 2 (Welfare)

Suppose $y_H - y_L \geq c$. As shown above, the agent provides sub-optimal effort in the no overtime equilibrium for any $s\nu\Delta < \frac{c-(y_H-y_L)(1-p)}{p}$. In this parameter range, which may also be empty, the principal receives the surplus $py_H + (1-p)y_L < S^*$ and the agent receives his outside option. For $c > s\nu\Delta \geq \frac{c-(y_H-y_L)(1-p)}{p}$, the efficient outcome obtains. The principal receives $y_H - c + s\nu p\Delta > py_H + (1-p)y_L$ and the agent a positive rent of $p(c - s\nu\Delta)$. For $s\nu\Delta \geq c$, the principal receives $S^*$. Hence, the principal's profit is at a maximum while the agent only receives his outside option payoff $s\bar{\theta}$.

Suppose $y_H - y_L < c$. If $s\nu\Delta \leq c$, then the principal implements the no overtime contract, which does not make any positive transfer to the agent, leaving the principal with the full surplus $S^*$. If $s\nu\Delta > c$, then the agent works too much. The principal has to compensate the agent for the cost of working overtime in expectation, so that the agent always receives his outside option payoff. The principal receives the remaining surplus which is less than the optimal surplus. Hence, the principal's profit is weakly decreasing in $s\nu\Delta$. Moreover, this would be true even if non-monotonic contracts were allowed. □

*Proof of Proposition 2.* We now allow contracts with transfers $W(y, \hat{e})$ that condition on reports. Note first that there is no pure strategy equilibrium $\hat{e}(\theta_L) = e_L$, $\hat{e}(\theta_H) = e_H$, $e(\theta_H) = e_L$ and $e(\theta_L) = e_H$ in which the low-skilled agent completes the task but reports are reversed such that overtime reports are associated with high skills and regular-time reports with low skills. This is obvious with $W(y_H, e_H) - W(y_H, e_L) = w \geq 0$, because then the high type would switch to overtime. Even without the monotonicity restrictions, reverse reports are excluded with our refinement: With lying costs, the incentive constraints for such reporting behavior are $W(y_H, e_L) - l_u \geq W(y_H, e_H) + s\nu\Delta$ and $W(y_H, e_H) + s\nu\Delta \geq W(y_H, e_L) + l_o$ and could never be satisfied for either $l_o > 0$ or $l_o > 0$. Hence, such an equilibrium would not satisfy our refinement.

Consider now the pure strategy equilibria discussed in the main text. Suppose $s\nu\Delta > c$. With revealed and false overtime the principal can maximally get $y_H - (1-p)s\nu\Delta$ which is strictly smaller than what she can get with hidden overtime, which is $y_H - (1-p)c$. Thus, out of the contracts that make the low type work overtime, the principal prefers the hidden overtime contract $(w^{HO}, b^{HO})$. With $s\nu\Delta > c$ the no overtime equilibrium cannot occur.

Suppose $s\nu\Delta \leq c$. In this case, the principal is indifferent between the contracts that induce overtime work $(w^{HO}, b^{HO})$, $(w^{FO}, b^{FO})$ or $(w^{RO}, b^{RO})$, which all yield profit $y_H - c + s\nu p\Delta$. The profit from this contract then only needs to be compared to the no overtime equilibrium, which is also implementable in this parameter range. This yields the conditions in the proposition.

Now consider the possibility of mixing between reports. If the agent produces low output, the type is already revealed by output. Following high output, since reports are cheap talk, there can for example be a "babbling" equilibrium, in which the types mix with equal probability among the two reports and reputation is always equal to the expected skill. However, our refinement rules out these equilibria, since with arbitrarily small lying costs it is no longer possible that the agent is indifferent between reports no matter what his type is. Another possibility are hybrid equilibria in which only one type mixes between reports while the other type reports actual effort. These equilibria can survive the refinement, since the type that tells the truth does not incur any lying costs.

If $s\nu\Delta \leq c$, the principal is indifferent between hidden, revealed, and false overtime, and he could also induce a mixed strategy. Mixing between revealed and false overtime would mean that the low type tells the truth and the high type mixes, mixing between revealed and hidden overtime would mean that the high type tells the truth and the low type mixes. If $s\nu\Delta > c$, the kind of mixing where the high type is indifferent between the two reports is never part of the optimal contract. It is possible, however, that the low type is indifferent between the reports. Since the agent receives no rent with hidden overtime, the same must be true if overtime is only partially hidden. Let the low type hide overtime with probability $\epsilon$. The indifference condition is $W(y_H, e_L) + s\nu\frac{p}{p+(1-p)\epsilon}\Delta = W(y_H, e_H)$ and the binding participation constraint is $W(y_H, e_H) = (1-p)c + s\nu p\Delta$. The condition $W(y_H, e_L) \geq 0$ gives the necessary condition $\epsilon \geq \frac{p(s\nu\Delta - c)}{p(s\nu\Delta - c) + c}$. Note that mixing between reports is never strictly optimal for the principal. □

*Proof of Proposition 3.* In Section 4.1, we have derived the least-cost contracts and have

shown that for positive lying costs, false overtime is always more expensive to induce than revealed overtime. The least-cost contract that induces

(i) hidden overtime exists if and only if $0 \leq s\nu p\Delta - l_u$ and yields profit

$$\pi^{HO} = \begin{cases} y_H - c - l_u + s\nu p\Delta & \text{if} \quad s\nu\Delta \leq c + l_u \\ y_H - (1-p)(c + l_u) & \text{if} \quad s\nu\Delta > c + l_u \end{cases}, \tag{26}$$

(ii) revealed overtime yields profit

$$\pi^{RO} = \begin{cases} y_H - c + p(s\nu\Delta + l_o) & \text{if} \quad s\nu\Delta \leq c - l_o \\ y_H - (1-p)c & \text{if} \quad c - l_o < s\nu\Delta \leq c + l_u \\ y_H - (1-p)(s\nu\Delta - l_u) & \text{if} \quad s\nu\Delta > c + l_u \end{cases}, \tag{27}$$

(iii) no overtime exists if and only if $s\nu\Delta \leq c + l_u$ and yields profit

$$\pi^{NO} = py_H + (1-p)y_L. \tag{28}$$

We consider next the possibility of mixing between reports. Lying costs make it impossible that both types are indifferent between the two reports. Consider first the case that the high type is indifferent between reports, which implies that the low type strictly prefers to tell the truth. A regular-time report is hence always attributed to the high type. The indifference condition for $0 < \mu(y_H, e_H) < p$ is

$$w + \mu(y_H, e_H)s\nu\Delta - l_o = s\nu\Delta, \tag{29}$$

and the incentive constraint for the low type to work overtime is

$$w + b + \mu(y_H, e_H)s\nu\Delta \geq c. \tag{30}$$

The agent's expected payoff in excess of $s(\nu\theta_L + (1-\nu)\bar{\theta})$ is $w + b + \mu(y_H, e_H)s\nu\Delta - (1-p)c - pl_o$. The participation constraint is therefore

$$w + b + \mu(y_H, e_H)s\nu\Delta - (1-p)c - pl_o \geq ps\nu\Delta, \tag{31}$$

and the total payment can also easily be obtained from this by adding the expected overtime and lying costs that are actually incurred by the agent and subtracting the average image benefit. Total payment is hence strictly larger than $w + b + \mu(y_H, e_H)s\nu\Delta - pl_o - ps\nu\Delta$. Constraints (29),(30), and (31) yield

$$w + b + \mu(y_H, e_H)s\nu\Delta \geq \max\{ps\nu\Delta + pl_o + (1-p)c, l_o + s\nu\Delta, c\},$$

where in the optimum this holds with equality. Because of the positive lying cost for a

completely mixed strategy of the high type, total payment is strictly larger than

$$\max\{(1-p)c, (1-p)(l_o + s\nu\Delta), c - p(s\nu\Delta + l_o)\},$$

which is weakly larger than the revealed overtime payment. Intuitively, the hybrid equilibrium is a mixture between false overtime and revealed overtime, and since revealed overtime dominates false overtime, it also follows that the hybrid equilibrium is dominated by revealed overtime.

Consider now the case that the low type is indifferent between reports. This time, the hybrid equilibrium is a mixture between hidden overtime and revealed overtime, and intuitively, it may have an advantage in saving lying costs compared to hidden overtime and an advantage of a lower rent compared to revealed overtime. For $p < \mu(y_H, e_L) < 1$, the indifference condition for the low type is

$$b + w = b + \mu(y_H, e_L)s\nu\Delta - l_u. \tag{32}$$

The low type exerts effort if

$$b + w \geq c. \tag{33}$$

The agent's expected payoff (in excess of $s(\nu\theta_L + (1-\nu)\bar{\theta})$) is $b + w + pl_u - (1-p)c$. Thus, the participation constraint is

$$b + w + pl_u - (1-p)c \geq ps\nu\Delta. \tag{34}$$

Denoting the probability with which the low type chooses to hide overtime by $\epsilon$, such that $\mu(y_H, e_L) = \frac{p}{p+(1-p)\epsilon}$, and using the indifference condition (32), total expected payment to the agent is equal to

$$b + (1-p)(1-\epsilon)w = b + w + pl_u - ps\nu\Delta + (1-p)\epsilon l_u,$$

where

$$b + w = \max\{c, \mu(y_H, e_L)s\nu\Delta - l_u, ps\nu\Delta - pl_u + (1-p)c\}.$$

Thus, total payment to the agent is

$$\max\{c - p(s\nu\Delta - l_u), (\mu(y_H, e_L) - p)s\nu\Delta - (1-p)l_u, (1-p)c\} + p(\frac{1}{\mu(y_H, e_L)} - 1)l_u.$$

If $s\nu\Delta - l_u \leq c$, this is dominated by revealed overtime. Else $\mu(y_H, e_L)$ can be chosen such that $(\mu(y_H, e_L) - p)s\nu\Delta = (1-p)(c + l_u)$. The agent, with total payment $(1-p)(c + \epsilon l_u)$, receives no rent, and compared to completely hidden overtime, some lying cost is saved. Note that lying costs can be saved even if they are small. Full hidden overtime is approached as $s\nu\Delta \to \infty$. To see that revealed overtime is indeed not implemented in this range, we find the optimal $\mu(y_H, e_L)$ by solving

$$\min_{\mu \geq p + (1-p)\frac{c+l_u}{s\nu\Delta}} \mu s\nu\Delta + \frac{p}{\mu}l_u.$$

It is straightforward to show that the objective function is increasing in the relevant range. Hence, the hybrid contract dominates revealed and hidden overtime in the parameter range $s\nu\Delta > c+l_u$, and the probability of hiding overtime solves $(\mu(y_H, e_L)-p)s\nu\Delta = (1-p)(c+l_u)$ which is equivalent to

$$\epsilon = \frac{p(s\nu\Delta - c - l_u)}{p(s\nu\Delta - c - l_u) + c + l_u}.$$

For $s\nu\Delta \leq c+l_u$, revealed overtime yields weakly highest profits among all contracts that induce overtime, and strictly so for positive lying costs. By Assumption 1, this is always better than having no contract at all.

Now we take into account also the no overtime contract and find the overall optimal contract. For $s\nu\Delta > c + l_u$, no overtime is not feasible and the hybrid contract described above is optimal. For $y_H - y_L \geq c$ and $c - l_o < s\nu\Delta \leq c + l_u$, the revealed overtime contract is optimal as it implements the first best. If $y_H - y_L < c$ and $s\nu\Delta \leq c + l_u$, the no overtime contract is optimal and implements the first best. For $y_H - y_L > c$ and $s\nu\Delta \leq c - l_o$, the revealed overtime contract is preferred to no overtime contract if and only if $(1 - p)(y_H - y_L) \geq c - p(s\nu\Delta + l_o)$. □

*Proof of Proposition 4.* To study the principal's incentives to install a recording system with lying costs $l_o, l_u \geq 0$, let $S(l_o, l_u)$ denote the surplus that is created with such a system by the contract characterized in Proposition 3.

**Case $s\nu\Delta \leq c + l_u$ and $y_H - y_L \leq c$.**

No overtime is efficient and implementable with the recording system, such that $S(l_o, l_u) = S^* \geq S(0, 0)$. It is efficient to install the system whenever $S^* - k_a - k_p > S(0, 0)$. Since the principal receives the full surplus before and after the recording system is installed, the principal installs it if and only if it is efficient to do so.

**Case $s\nu\Delta > c + l_u$ and case $c \leq s\nu\Delta \leq c + l_u$ and $y_H - y_L \geq c$.**

Without a recording system, the principal implements hidden overtime and receives the full surplus that is generated. Installing the system also leads to overtime but some costs might be incurred. Installing the system therefore never increases the generated surplus in these cases. The principal will therefore not install a recording system, and this is also the efficient choice.

**Case $s\nu\Delta < c$ and $(1 - p)(y_H - y_L - c) > p(c - s\nu\Delta)$.**

Without the recording system, the agent receives a rent, which after installing a recording system with positive $l_o > 0$ or $k_A > 0$ decreases to $\max\{p(c - s\nu\Delta - l_o) - k_A, 0\}$. This implies that the principal will install the recording system whenever $\min\{pl_o, p(c - s\nu\Delta) - k_A\} > k_P$. Since the surplus stays the same, the efficient choice is not to install the system, and this is also what the principal chooses for $l_o = 0$, which is assumed in the proposition. Only if $l_o > 0$, rent-seeking by the principal could lead to an inefficient adoption of a recording system.

**Case $s\nu\Delta < c$ and $0 \leq (1-p)(y_H - y_L - c) \leq p(c - s\nu\Delta)$.**

The no overtime contract obtains without a recording system. Any recording system that does not change this outcome is inefficient and would not be installed by the principal. Consider now the remaining case that there is a switch from the no overtime contract, which leaves the agent without a rent, to the revealed overtime contract, which might pay a rent to the agent. The agent's rent is given by $A^* = \max\{p(c - s\nu\Delta - l_o) - k_A, 0\}$. The recording system increases the surplus by $(1-p)(y_H - y_L - c) - k_P - k_A$, but the principal's profit only by $(1-p)(y_H - y_L - c) - A^* - k_P - k_A$. Hence, there is a parameter constellation here in which the principal does not install the system although it would be efficient to do so and the agent would benefit. However, only recording systems with $l_o > 0$ may induce a switch from no overtime to revealed overtime. □

*Proof of Proposition 5.* Let any recording system $r = (l_u, l_o, k_A, k_P)$ be described by lying costs $l_u, l_o$, reporting cost $k_A$ and a set-up cost $k_P$. Suppose there is a set $R$ of recording systems available. Let $(0,0,0,0) \in R$ denote the possibility to not install any recording system. For any given system $r$, there is an optimal contract and equilibrium payoff as described in Section 4.2. Let $S(r)$ denote the attained surplus, and let $A(r)$ be the agent's rent (expected payoff minus average image utility $s\bar{\theta}$). Let furthermore $\hat{A} = \max_{r \in R} A(r)$. Finally, let $R^*$ denote the set of optimal recording systems, i.e. those that maximize the surplus including all transaction costs $S(r) - k_A(r) - k_P(r)$.

The agent can never receive more than $\hat{A}$. If $\hat{A} > 0$, then this maximum rent is obtained by a recording system with $l_o > 0$, meaning that the agent can require this system and thereby guarantee himself $\hat{A}$. Since the agent can always reject the contract, he is also guaranteed to receive at least a payoff of zero. If the principal installs an optimal system $r^* \in R^*$ and offers the optimal contract for this recording system with an additional payment of $\hat{A} - A(r^*)$ to the agent, then the agent will not change the system. Hence, the principal can guarantee herself the payoff $S(r^*) - k_P(r^*) - k_A(r^*) - \hat{A}$. Since the agent can guarantee himself $\hat{A}$ and the principal can guarantee herself the maximum joint surplus minus $\hat{A}$, this must be the outcome of any equilibrium. Intuitively, the limited liability of the agent is a transaction cost that precludes efficient Coasean bargaining, and therefore giving more property rights to the wealth-constrained agent can restore efficiency. □

# B Interpretations of the Model

Here we illustrate how self-deception may be facilitated by motivated effort denial and how the image utility can be interpreted as stemming from career concerns.

## B.1 Motivated Beliefs

In the model in the main part of the paper, effort denial is something that occurs naturally and is not influenced by the agent. However, *motivated beliefs* might play a role here, meaning that an agent may strategically choose what to forget or even strategically choose beliefs. Therefore we now consider a model of motivated beliefs as in Bénabou and Tirole (2002) and

Bénabou (2015) with a Bayesian agent. The difference to the model in the main part of the paper is that the agent voluntarily, or strategically, forgets effort. Skill is always forgotten, or anyway only inferred from the project's outcome after a regular amount of work - in this sense, remembering skill is tied to remembering effort.

The agent has two selves. The date-1-self makes the decision to work and report and also encodes a memory of exerted effort, which means that the date-1-self sends a signal about effort to his date-2-self. The agent can either opt for realism and send a true signal at zero cost, or distort information and send the false signal at cognitive cost $m \geq 0$. This cost $m$ shall be small, in particular smaller than the maximum gain from image concerns, $m < s\Delta$. The date-2-self knows that his memory cannot be completely trusted and makes an update relying on output and report in addition to the signal. The date-2-self does not choose an action but merely consumes these beliefs.

In our framework, the agent's beliefs still affect none of his decisions and therefore have no instrumental value. Since the agent prefers to believe to be high-skilled, it is never optimal for the high-skilled type to fabricate a memory of having worked overtime. As long as belief distortion has no direct cost, remembering low effort is also a weakly dominant strategy for the low type. When both types pool on the same observable behavior, the updated belief is then equal to the prior. As a consequence, for $m = 0$ this model with motivated beliefs is equal to the model in the main part of the paper for the special case $\nu = 1$. While this already shows how the model in the main part can be interpreted as a model of motivated beliefs, the following paragraphs sketch the extension to the case $m > 0$ and show that the main results also basically stay the same.

The cost of cognitive distortion $m$ acts like a lying cost that is incurred only by the low type and irrespective of whether the principal asks for a report or not. The conditions for hidden overtime stay the same with $m$ added to $l_u$. The same is true for no overtime, meaning that no overtime is now implementable in a slightly larger range of parameters. Moreover, no overtime has the advantage that the cost of cognitive distortion is not incurred.

The conditions for revealed overtime with motivated beliefs are also similar. Again, $m$ is added to $l_u$. The overtime payment that reveals overtime now has to satisfy

$$s\Delta + l_o - m \geq w \geq s\Delta - l_u - m.$$

Here it is important that it is structurally possible in this model to remember low effort while reported effort was high. In the out-of-equilibrium event that the agent remembers timely completion but has reported overtime, he does not know whether to trust his memory or his report. His belief in this case is pinned down by the refinement. The low type benefits more from the deviation to forget overtime work than the high type benefits from the deviation to report overtime if $s\Delta + l_o - w - m \geq 0$. This condition is needed for truth-telling to be an equilibrium that survives our refinement, and it is stricter than the high type's incentive constraint $s\Delta \geq w - l_o$.

With false overtime, the low type has to incur the cost $m$ if he wants to profit from the better self-image. This cost makes false overtime even less attractive and in fact, false overtime is now strictly dominated by revealed overtime as in the case of positive cost of

claiming overtime $l_o > 0$.

By comparing these cases, one can again find the optimal contract for the principal. For example in the case without lying costs this optimal contract is as follows:

**Corollary 4.** *For $s\Delta \leq c+m$, the revealed overtime contract is optimal whenever $(1-p)(y_H - y_L - c) \geq p(c+m-s\Delta)$, else no overtime is optimal. For $s\Delta > c+m$, the optimal contract induces overtime work that remains hidden with probability $\frac{p(s\Delta-c-m)}{p(s\Delta-c-m)+c+m}$.*

This shows that the difference between selective and involuntary forgetting is not large. The cost of avoiding to accurately recall high effort acts as a tie-breaker between hidden overtime and revealed overtime in the case without lying costs and is otherwise simply added to the costs of working overtime.

Larger departures from the model in the main part of the paper are possible for example by assuming that the agent selects what to remember before he makes the report. Moreover, with the interpretation of self-deception, the assumption that the principal can choose the equilibrium is not as natural anymore, and one might consider allowing the agent to choose the equilibrium. This assumption would affect the results, since for a given overtime payment, the agent prefers false overtime to revealed overtime whenever both are equilibria that survive the refinement.

## B.2 Career Concerns

If the agent is concerned about his reputation because of career concerns, it is in fact very likely that the principal also cares about learning the type, and that therefore revealed overtime will be chosen more often. To stay close to the model in the main part of the paper, we present here a simple career concerns model that does not have this feature because the principal never fires the agent, even if she learns that he is a low-skilled type.

Assume that the agent and the principal live for two periods. During the first period, working for the principal lets the agent acquire relationship-specific knowledge such that the situation in the second period resembles a bilateral monopoly, with no other agent being able to create a comparably high surplus for the principal. To model career concerns, we assume that in the second period, output depends only on skill and is $\theta_H$ for the high type and $\theta_L$ for the low type. The productivity gain from relationship-specific experience is assumed to be so large that the low type's second-period output is larger than the high type's first-period output, $\theta_L \geq y_H$.

After the first period, agent and principal have to agree on the terms for the second period.[16] The agent can leave the relationship and receive an outside payoff of zero. We assume that at the start of the second period, the agent makes a take-it-or-leave-it offer to the principal with probability $s$ and the principal makes a take-it-or-leave-it offer to the agent with probability $1-s$. If the agent rejects the principal's offer, the principal hires a new agent, pays him zero and receives $\bar{y} = (1-p)y_L + py_H$.

In this negotiation, the agent accepts any non-negative wage offer from the principal, which means that he will receive a wage of zero if the principal gets the right to make the

[16] We thus assume that long-term contracts are not possible. What is important here, however, is not that there can be no long-term contract but that the parties cannot commit to not renegotiate such a contract.

offer. The principal accepts a wage offer from the agent if the wage is not higher than the expected productivity in excess of $\bar{y}$. Hence, if the agent makes the offer, he demands a wage of $E[\theta|y,\hat{e}]-\bar{y}$. Overall, the agent's expected utility from the second period is $s(E[\theta|y,\hat{e}]-\bar{y})$. If the agent's skill type is $\theta$, the principal's expected profit in the second period is $(1-s)\theta+s(\theta-E[\theta|y,\hat{e}]+\bar{y})$. Since the principal always keeps the agent employed for the second period, she does not benefit from learning his type. The law of iterated expectations implies that seen from the start of the first period, the principal receives an expected payoff of $(1-s)\bar{\theta}+s\bar{y}$ in the second period. The principal's expected second-period payoff hence acts like a constant.

We assume here that as an outside option in the first period, the agent could work for another principal, where the task has a simpler structure without the possibility of working harder to complete the task, and where he could also acquire relationship-specific knowledge and receive $s(\bar{\theta}-\bar{y})$ in the second period. With these assumptions, the analysis is exactly the same as in the main part of the paper.