

Inconsistent Time Preferences in the Labor Market

– When it Pays to be Naive*

Matthias Fahn[†] & Regina Seibel[‡]

September 18, 2021

Abstract

We study optimal employment contracts for present-biased employees if firms cannot commit to long-term contracts. Assuming that an agent's effort increases his chances to obtain a future benefit, we show that individuals who are naive about their present bias will actually be better off than sophisticated or time-consistent individuals. Moreover, firms might benefit from being ignorant of the extent of an employee's naiveté. Our results also indicate that naive employees might be harmed by policies such as employment protection or a minimum wage, whereas sophisticated employees are better off.

JEL Codes: D21, D90, J31, J32

Keywords: Present bias, labor markets, on-the-job search, moral hazard

*We are grateful to Florian Ederer, Florian Englmaier, Paul Heidhues, Takeshi Murooka, Nick Netzer, Armin Schmutzler, Heiner Schumacher, Peter Schwardmann, Ferdinand von Siemens, as well as seminar participants at the COPE 2018 (Munich) and the annual meeting of the Verein fÄr Socialpolitik 2018 (Freiburg) for valuable comments and suggestions. Financial support by Deutsche Forschungsgemeinschaft through CRC TRR 190 is gratefully acknowledged.

[†]JKU Linz & CESifo; matthias.fahn@jku.at

[‡]University of Zurich; regina.seibel@econ.uzh.ch

1 Introduction

People suffer from self-control problems which are often caused by inconsistent time preferences. A huge literature has explored how firms lure consumers into inefficient “exploitative contracts” and thereby extract substantial rents from those who are naive about their present bias.¹ However, although recent observations suggest that time-inconsistent preferences matter in the workplace as well (Kaur et al., 2010, 2015), evidence that firms also try to attract and exploit naive *employees* remains scarce.² Naturally, a limited amount of evidence does not implicate the absence of exploitative contracts in the workplace. Nonetheless, wrong expectations concerning one’s future behavior might have different consequences for employees than for consumers because employment relationships are inherently incomplete and often lack (formal) long-term commitment.

In this paper, we show that misperceptions of their future behavior can benefit employed individuals if firms cannot commit to long-term contracts. As a consequence, firms’ profits are lower when hiring individuals they know to be naive. Moreover, being ignorant about an employee’s naiveté can *increase* a firm’s profits. Otherwise, firms might completely abstain from hiring naive employees.

We derive these results within a three-period model in which a principal can hire an agent. In periods 1 and 2, the principal can make a take-it-or-leave-it offer to the agent but is not able to commit to long-term contracts. Upon acceptance, the agent chooses his effort level, while a rejection ends the game. Higher effort is associated with higher costs for the agent and increases the likelihood of receiving some benefit in the *subsequent* period. We first assume this benefit to be exogenously given. Later, we consider two applications. In the first application, effort reflects on-the-job search and the benefit is an outside job offer. In the second application, effort increases the principal’s profits, and the agent’s benefit is a performance-based bonus as in a standard moral hazard case. Importantly, the chance to exert effort is tied to the job, thus the possibility to collect the benefit can be regarded as a non-pecuniary advantage of employment. This allows the principal to pay a negative wage premium that pushes the agent’s compensation below his outside option of zero.

The agent has time-inconsistent preferences and is present biased, thus his effort is lower than the effort of a time-consistent agent with the same long-run discount factor. This implies that also the surplus generated by a time-inconsistent agent is smaller. Since the principal has full bargaining power, the offered wage

¹See DellaVigna (2009), Kőszegi (2014), or Heidhues and Kőszegi (2018) for overviews.

²Exceptions are Bubb and Warren (2020) and Hoffman and Burks (2020).

completely extracts the rent from effort exertion that is generated in a given period. Moreover, an agent's time inconsistency affects how he perceives the rents caused by *future* effort: Discounting between the second and the third period is larger from the perspective of period 2 than from the perspective of period 1. Thus, the rent of period-2 effort is larger if assessed from earlier periods. In other words, a time-inconsistent agent enjoys an "extra rent" of his future effort.

Now, a time-inconsistent agent can either be sophisticated or naive (Laibson, 1997, O'Donoghue and Rabin, 1999b).³ Whereas the sophisticated agent perfectly anticipates his future present bias, the naive agent expects to be time-consistent later on. Therefore, while second-period effort and wage are the same for naive and sophisticated agent, the naive agent misperceives his values when assessing future outcomes in period 1. He overestimates his second-period effort and thus underestimates the wage he is going to accept. Moreover, he does not anticipate his extra rent from period-2 effort. Therefore, his first-period reservation wage only incorporates the rent created by this period's effort. This is different for the sophisticated agent who is aware of his future time inconsistency and thus takes the extra rent of period-2 effort into account. Because this extra rent is tied to employment, it lets him accept a lower first-period wage than the naive agent. All this implies that the long-run utility (i.e., the utility evaluated in an imaginary period before the game starts) of a naive agent exceeds the utility of a sophisticated agent. This outcome relies on the principal's lack of commitment, and additionally on our assumption that the game ends if the agent rejects an offer.⁴ Note that, although the sophisticated agent's long run utility is below the naive agent's, it still exceeds the level of a time-consistent agent. The reason is that also period-1 effort generates some extra utility if assessed in the imaginary pre-game period, and this extra utility cannot be exploited by the principal.

In the next steps, we consider two applications that allow us to generate additional insights on the consequences of present bias in the labor market. First, we interpret effort as on-the-job search and the benefit as an outside offer.⁵ Different from the benchmark model, successful search does not generate a one-time benefit but permanently increases the agent's outside option because of a better visible

³In our main specification, we assume that the principal can observe the extent of the agent's present bias as well as naiveté. We relax this assumption in Section 6.3 and allow for uncertainty about the extent of naiveté, as in Eliaz and Spiegel (2006).

⁴If a first-period rejection still entailed a second-period offer, the sophisticated agent's utility would be higher. In Section 8, we discuss a potential microfoundation for our assumption.

⁵Our assumption that effort can only be exerted while employed requires on-the-job search to be more effective than search out of unemployment. Indeed, Biewen and Steffes (2010), Mueller (2010), or Cingano and Rosolia (2012) present evidence that this is the case.

position on the labor market.⁶ As before, the naive agent underestimates his second-period wage, which drives up his first-period reservation wage and increases his long-run utility above the level of the sophisticated agent. In addition, he searches more than the sophisticated agent. The reason is that the future rent caused by the agent’s time inconsistency increases in future effort, which is higher after search has not been successful. Since only the sophisticated agent takes this link into account, his perceived utility of *not* receiving an offer is higher, thus his first-period search effort smaller.

Second, we assume that effort benefits the principal by increasing the probability with which some output is generated. Only this output is observable and verifiable, thus the principal can set a performance-based bonus. Then, the naive agent overestimates his second-period bonus and underestimates his second-period wage, thus his first-period reservation wage and his long-run utility again exceed the levels of the sophisticated agent. Besides the first-period wage difference, the realized contracts and effort choices of naive and sophisticated agent turn out to be identical, a result which is the same as in the baseline model. Therefore, the principal’s profit with a sophisticated agent is higher than with a naive agent.

Our results stand in stark contrast to most of the literature on present-biased preferences, where naive individuals generally are worse off. Firms design “exploitative contracts” to actively attract consumers they expect to mispredict their own future use of a purchased product and charge high prices when agents change their plans.⁷ Our paper indicates that these results cannot necessarily be extended to employment relationships if firms are not able to commit to long-term contracts because the naive agent’s pessimism about his future wage prevents him from accepting low wages in the present.

In a number of extensions, we explore variations which affect the naive agent’s misperception of his future utility. Importantly, his higher compensation in period 1 is caused by the principal’s inability to commit to the “real” period-2 wage which exceeds the agent’s perceived wage. Thus, we first allow for (partial) commitment by the principal. We assume that she can announce a second-period contract already in period 1, but is then able to terminate the relationship at some cost. Higher separation costs are therefore equivalent to more commitment and allow the principal to credibly offer a second-period wage that exceeds the amount the naive agent would otherwise expect. This allows for a reduction of the naive agent’s first-period wage and consequently his long-run utility. If separation costs are sufficiently high,

⁶Note that we allow for both possibilities, that the agent moves to another employer or that he is retained by the principal.

⁷See Heidhues and Kőszegi, 2010, or Kőszegi, 2014 for a survey.

the naive agent might even be worse off than the sophisticated agent because he overestimates his second-period effort and thus his second-period rent, letting him accept an even larger wage reduction in the first period. We argue that higher separation costs are reminiscent of more stringent employment protection, which therefore might harm the naive agent.

A further interaction of labor market policies with the trade-offs generated by our model is explored in the second extension, where we allow for a minimum wage. We show that a minimum wage can harm the naive agent if it is below the payment he ends up receiving, but above the (lower) wage he expects to be paid in the future. Then, the minimum wage leads to a reduction of the agent's misperception, which again allows the principal to lower the wage in the first period. This result indicates that a non-binding minimum wage might *reduce* wages; indeed, evidence for such negative spillover effects has been presented by Neumark et al. (2004), Stewart (2012), or Hirsch et al. (2015). Furthermore, a minimum wage always makes the sophisticated and time-consistent agent (weakly) better off.

In a final extension, we consider asymmetric information in the sense that the principal is not able to observe whether the agent is naive or sophisticated. As Eliaz and Spiegel (2006), we assume that the agent's present bias is common knowledge. Different from previous research, we find that the principal's profits when being ignorant can actually be larger than if she can observe the agent's naiveté. Hence, personality tests in the hiring process which allow firms to develop a better idea about their employees' characteristics might have negative side effects. The exact outcome depends on the naive agent's inter-player perceptions, i.e., whether he perceives other agents to be time-consistent in the future or not. In the latter case which has received more empirical support (Fedyk, 2018), the principal benefits from asymmetric information. Then, the naive agent anticipates a higher second-period wage offer than under symmetric information and consequently accepts a lower wage in the first period.

Our paper is organized as follows. We provide a short literature review in Section 2. The theoretical setup for the baseline model is described in detail in Section 3, which is then followed by a statement and intuition of our main results in Section 4. We turn to an analysis of our two applications in Section 5, and cover the extensions in Section 6. Finally, in Section 7 we show that our results are robust to allowing for partial naiveté, discounting between wage payment and effort choice, and extending the time horizon. We conclude in Section 8.

2 Literature Review

Contracting with present biased and potentially naive agents has been analyzed in the context of consumption as well as labor contracts.⁸ In the latter category, most papers consider moral hazard problems in principal-agent models, similar to our application in Section 5.2. A seminal contribution by O’Donoghue and Rabin (1999b) has considered the optimal incentive scheme for the completion of a single task. While similar in spirit, Yılmaz (2013) considers effort choice and compares a sophisticated with a time-consistent agent. Moreover, it is a consistent finding that naiveté is detrimental for the agent. Englmaier et al. (2016) show that the optimal exploitative menu of contracts for the principal consists of a virtual contract which the naive agent intends to choose in the future, and a real contract which he ends up choosing. The harm for the naive agent also occurs in models with imperfect information on the side of the principal, such as Eliaz and Spiegler (2006) and Gilpatric (2008), where the optimal contract exploits naive, but not sophisticated, agents. In Eliaz and Spiegler (2006), the principal can forecast the behavior of the present-biased agents, but does not know the extent of naivete that the agent exhibits when forecasting his own behavior.⁹

Exploitative contracts have also been identified in interactions between consumers and firms, e.g., by Heidhues and Kőszegi (2010), Heidhues and Kőszegi (2017) and Gottlieb and Zhang (2021). Gottlieb and Zhang (2021) show that optimal consumption contracts do not depend on the consumer’s extent of naiveté, whose harm vanishes as the number of periods goes to infinity. Moreover, they consider contracting on consumption without an unobserved effort component. More generally, previous papers assume full commitment by the firm (or principal). In contrast, we consider a principal-agent setting where the principal may not be able to commit to future labor contracts. Then, a naive agent is better off than a time-consistent agent, a relationship which holds for any number of periods and does not vanish as the number of periods goes to infinity.

Only a few papers consider the effect of inconsistent time preferences on job search behavior. However, a large numbers of job-to-job transitions indicate that on-the-job search is a significant force behind labor market dynamics. For example, Bjelland et al. (2011) find that employer-to-employer flows accounted for around 4 percent of total employment in the US between 1991 and 2003; see Fallick and Fleischman (2001) or Nagypál (2008) for further evidence. Search activities on labor

⁸See Kőszegi (2014) and Grubb (2015) for reviews.

⁹We apply this assumption in our asymmetric information setting in section 6.3. Since the principal cannot condition compensation on effort, however, she cannot screen the agent as described in Eliaz and Spiegler (2006) or Gilpatric (2008).

markets are mostly perceived to be caused by information frictions which prevent an immediate matching of workers with their optimal job types. There, heterogeneities of workers and jobs have gained considerable attention as main drivers of these frictions, see Pissarides (1994), Mortensen (2000), Moscarini (2005), or Gautier et al. (2010). But less focus has been put on how the trade-off between costly search effort today and potential benefits later on determines the extent, and consequently stickiness, of the generated inefficiencies. Exceptions are DellaVigna and Paserman (2005) and Paserman (2008) who, among others, have recently incorporated “behavioral” assumptions into job search models. They show that more present biased agents search less and set lower reservation wages. In our application in Section 5.1 we derive similar results for on-the job search.

We also contribute to the literature on the relationship between present bias and commitment, which however mostly focuses on the sophisticated agent’s ability to commit. The preference description by O’Donoghue and Rabin (1999a) of the present-biased but sophisticated agent reveals that he might benefit from limiting his future set of feasible choices. More recently, Amador et al. (2006) and Bond and Sigurdsson (2018) have analyzed contracts which might help the agent by offering commitment. Kaur et al. (2015) model and show empirically that the demand for commitment in a labor market context is indeed affecting choices of employees, leading them to choose steep incentives for themselves. We complement these approaches and analyze commitment by the principal, not the agent. Our results imply that commitment benefits the principal as well as the sophisticated agent. Naive agents, however, are harmed by a stronger commitment of the principal.

3 Baseline Model

Environment, Technology & Contracts

There is one principal (“she”) and one agent (“he”), we analyze a game with three periods, $t = 1, 2, 3$.¹⁰ At the beginning of the first and the second period, the principal can make a take-it-or-leave-it employment offer to the agent, which consists of an upfront payment $w_t \in \mathbb{R}$. If the agent rejects the offer, the game ends and the agent consumes his outside option which is normalized to zero. Upon acceptance, the agent chooses an effort level $e_t \in [0, 1]$ which is associated with effort costs $e^2/2$. Moreover, e_t equals the probability with which the agent receives some benefit $b > 0$ *at the beginning of the next period*. Thus, the possibility to exert effort and

¹⁰Results are robust to considering a large finite number of periods as well as an infinite time horizon; see Section 7.3.

potentially receive the benefit are tied to employment. Moreover, players are not active in the third period, only the agent potentially receives the benefit b .

We allow for several interpretations of the agent’s effort associated with different specifications of the benefit b . First, in the simplest setting, the agent’s effort leads to an exogenous one-time benefit which either does not benefit the principal directly or cannot be incentivized. We might interpret this as effort for the accomplishment of private projects, for which the agent requires resources or reputation only the employer can provide. An example would be a researcher who is intrinsically motivated to conduct some research project, but who needs data provided by the employer. The benefit of a successful project is just a personal satisfaction which the employer cannot directly influence. This is the setting we introduce formally in the current section and analyze in Section 3.

Second, a successful project could lead to better job opportunities, inducing us to interpret effort as search for a better job (“effort as on-the-job search”). This interpretation is different because a success changes the agent’s utility permanently: He will either switch to a better employer or the current employer has to give him a raise in order to keep him. This setting is explored in more detail in Section 5.1. Third, in Section 5.2, we show that our results also apply to a more standard moral hazard setting, where the employer benefits from the employee’s effort, thus wants to provide incentives and sets a bonus endogenously (“effort as a task that benefits the principal”).

Contracting Effort e_t is the agent’s private information and consequently not contractible. Generally, employing the agent has some inherent value for the principal. For example, the agent might conduct other tasks which are taken care of by a (not further modelled) incentive contract. In such a case, w_t would correspond to the agent’s net utility from employment (besides effort costs and the benefit b). For simplicity, though, we stick to using the term wage when referring to w_t .

Crucially, the principal can only offer *short-term employment contracts*. In particular, in the first period she is not able to commit to any second-period wage. We discuss the importance of this assumption and the consequences of allowing for (partial) commitment in Section 6.1. If the agent is employed in the first period, the benefit from e_1 is not tied to employment in period 2.

Preferences

The agent is risk neutral and discounts future costs and future utilities in a quasi-hyperbolic way according to Laibson (1997) and O’Donoghue and Rabin (1999a). Immediate utilities are not discounted. Utilities at the next stage of a period are

discounted with a factor $\beta\delta$ and utilities after t periods are discounted with a factor $\beta\delta^t$, where $\beta \in (0, 1]$. We normalize δ to 1 as it has no qualitative effect on our results.

Hence, an agent's preferences may be dynamically inconsistent. This implies that, conditional on accepting the principal's offers, the agent's utility at the beginning of period $t = 1$ equals

$$U_1 = w_1 - \frac{1}{2}e_1^2 + \beta \left\{ e_1b + \left[w_2 - \frac{1}{2}e_2^2 + e_2b \right] \right\}.$$

The agent's utility at the beginning of period $t = 2$ equals

$$U_2 = w_2 - \frac{1}{2}e_2^2 + \beta \{e_2b\}.$$

A comparison between U_1 and U_2 reveals the agent's time inconsistency. Whereas there is no discounting between periods 2 and 3 from the perspective of period 1, the effective discount factor falls to β if evaluated from the perspective of period 2.¹¹

The principal is not present biased. Since the bonus is exogenous in the baseline setting, it is sufficient to consider the minimal wage still accepted by the agent without further specifying the principal's profits. We describe the principal's problem in more detail for the case that effort directly benefits the principal and b is a performance-based bonus (Section 5.2).

Perceptions

We assume that the agent might be sophisticated or (fully) naive concerning his future present bias.¹² A naive agent expects his present bias to disappear and to discount the future exponentially from the next period on. In contrast, a sophisticated agent perfectly anticipates his future present bias and thus also his future behavior.

Concerning inter-player perceptions, we assume common knowledge about the principal's time preferences. Moreover, the principal is aware of the agent's present bias as well as whether he is naive or sophisticated.¹³ However, whereas the principal anticipates potential contradictions between planned and realized actions, the agent thinks that the principal shares his own perception regarding his future preferences.

¹¹In Section 7.2 we discuss the implications of the present bias referring to *all* subsequent actions or outcomes. Then, upon receiving the period- t wage, the agent would already discount this period's search effort with β .

¹²In Section 7.1, we explore partial naiveté.

¹³In Section 6.3, we assume that the principal cannot observe the extent of the agent's naiveté.

Equilibrium

Following O'Donoghue and Rabin (1999a) and Englmaier et al. (2016), our equilibrium concept is perception-perfect equilibrium. There, a player's strategy maximizes expected payoffs in all subgames, given one's present preferences, and given one's perceptions of one's own future behavior as well as of the others'. This equilibrium concept enables us to support strategies that are built on a naive agent's inconsistent beliefs.

4 Results

In the following, we solve for a perception-perfect equilibrium that maximizes the principal's profits. Since the principal cannot commit to long-term contracts, her profits are maximized at the beginning of every period, and we have to apply backwards induction to solve for equilibrium outcomes. In a profit-maximizing equilibrium, wage payments are minimized in every period. We will start with the time-consistent agent as a benchmark case for optimal effort and wage setting. We then characterize equilibria for sophisticated and fully naive agents separately and subsequently compare the outcomes.

Benchmark: Time-Consistent Agent

In the second period, conditional on having accepted the principal's employment offer, the time-consistent agent chooses effort to maximize $-e_2^2/2 + e_2b$, which yields an effort level

$$e_2^{TC} = b.$$

Since the total utility of exerting effort, $-(e_2^{TC})^2/2 + e_2^{TC}b = b^2/2$, is strictly positive, and since the agent can only exert effort if he is employed by the principal, the period-2 reservation and thus offered wage equals

$$w_2^{TC} = \frac{1}{2}(e_2^{TC})^2 - e_2^{TC}b = -\frac{1}{2}b^2 < 0.$$

Taking into account w_2^{TC} and expected net benefits of effort exertion, period-2 utility equals the agent's reservation utility of zero. Hence, the situation in the first period is equivalent to the second period, which implies that outcomes coincide as well. Lemma 1 collects the results for the case of a time-consistent agent.

Lemma 1 *A time consistent agent*

- *exerts the same effort in periods 1 and 2, i.e. $e_1^{TC} = e_2^{TC}$*

- receives the same wage in periods 1 and 2, i.e. $w_1^{TC} = w_2^{TC} < 0$.

The proof can be found in Appendix A.1.

4.1 Sophisticated Agent

Now, we analyze outcomes for a present biased but sophisticated agent. In the second period, having accepted the principal's employment offer he chooses effort to maximize $-e_2^2/2 + \beta e_2 b$, which yields an effort level

$$e_2^S = \beta b.$$

As with the time-consistent agent, the period-2 wage w_2^S takes into account that the agent can only exert effort and subsequently collect b if he is employed, thus is set to satisfy $U_2^S = w_2^S - (e_2^S)^2/2 + \beta e_2^S b = 0$. Therefore,

$$w_2^S = \frac{1}{2}(e_2^S)^2 - \beta e_2^S b = -\frac{1}{2}(\beta b)^2 < 0.$$

The principal can reduce the wage below the agent's outside option and extract the agent's rent from his effort. Moreover, the agent's time-inconsistency gives the principal additional, intertemporal, opportunities to reduce wages.

Lemma 2 *Assume the agent is sophisticated. Then,*

- effort in the first period is the same as in the second period, i.e., $e_1^S = e_2^S$
- the period-1 wage is lower than the period-2 wage, i.e., $w_1^S = w_2^S - (1 - \beta)(\beta b)^2 < w_2^S$.

The proof can be found in Appendix A.2.

From the perspective of period 1, there is no discounting between periods 2 and 3, whereas the potential third-period benefit b is discounted with β from the perspective of period 2. This changes the relative assessment of costs and benefits of period-2 effort. Thus, although w_2^S fully extracts the agent's net utility from effort in period 2, it does so only *from the perspective of period 2*. From the perspective of period 1, however, the agent's period-2 net utility from effort is higher. Plugging $w_2^S = \frac{1}{2}(e_2^S)^2 - \beta e_2^S b$ into the agent's period-1 utility yields

$$U_1^S = w_1 - \frac{1}{2}(e_1)^2 + \beta [e_1 b + \beta(1 - \beta)b^2].$$

There, the last term, $\beta(1 - \beta)b^2$, captures the “extra” utility of period-2 effort when assessed from the perspective of earlier periods.

Therefore, the wage w_1^S not only extracts the agent’s rent from period-1 effort, but also the agent’s extra benefits from period-2 effort. Finally, note that, from the perspective of period 1, the agent exerts too little effort for his own taste in period 2 (b versus βb).

4.2 Naive Agent

Now, we assume that the agent is naive about his present bias which implies that, in period 1, he wrongly believes to be time-consistent in the second period. Therefore, we have to distinguish between his realized and his anticipated second-period effort.

Having accepted the principal’s employment offer, the naive agent’s realized effort in period $t = 2$ also maximizes $-(e_2)^2/2 + \beta e_2 b$, yielding an effort level

$$e_2^N = \beta b.$$

Furthermore,

$$w_2^N = \frac{1}{2}(e_2^N)^2 - \beta e_2^N b = -\frac{1}{2}(\beta b)^2 < 0,$$

thus $w_2^N = w_2^S$ and $e_2^N = e_2^S$. However, from the perspective of period 1 the agent anticipates to maximize $-\frac{1}{2}(e_2)^2 + e_2 b$ and choose an effort level

$$\tilde{e}_2^N = b.$$

Because $\tilde{e}_2^N > e_2^N$, the agent overestimates his future effort. As a consequence, in period 1 the naive agent *underestimates* his period-2 wage. He expects to be offered a wage $\tilde{w}_2^N = (\tilde{e}_2^N)^2/2 - \tilde{e}_2^N b = -(1/2)b^2$ which is smaller than the period-2 wage he is effectively willing to accept, w_2^N .

The naive agent’s behavior in $t = 1$ is thus determined by his perceptions of future outcomes, not their true realizations:

Lemma 3 *Assume the agent is naive. Then,*

- *efforts in the first and second period are equal, i.e., $e_1^N = e_2^N$*
- *the period-1 wage is equal to the period-2 wage, i.e., $w_1^N = w_2^N$.*

The proof can be found in Appendix A.3.

From the perspective of period $t = 1$, the naive agent perceives his period-2 net utility to be zero. The principal thus is not able to collect the additional effort benefits that stem from the agent’s future time inconsistency.

4.3 Comparison

Finally, we compare outcomes of a time-consistent, a naive, and a sophisticated agent. First, recall that $e_2^S = e_2^N < e_2^{TC}$ and $w_2^S = w_2^N > w_2^{TC}$. Therefore, realized outcomes in period 2 are identical for a sophisticated and a naive agent, but both exert less effort and receive a higher wage than a time-consistent agent.¹⁴ However, the naive agent expects to exert the same effort and receive the same wage as the time-consistent agent, i.e., $e_2^S < \tilde{e}_2^N = e_2^{TC}$ and $w_2^S > \tilde{w}_2^N = w_2^{TC}$. This lets period-1 wages of a naive and a sophisticated agent differ.

Proposition 1 *In the first period,*

- *effort of a naive agent and a sophisticated agent are the same, and both are lower than the effort of a time-consistent agent, i.e., $e_1^S = e_1^N < e_1^{TC}$*
- *the wage of the naive agent is higher than the wage of a sophisticated agent, which in turn is higher than the wage of a time-consistent agent, i.e., $w_1^{TC} < w_1^S < w_1^N$.*

The proof can be found in Appendix A.4.

From the perspective of period 1, a sophisticated agent perceives his period-2 net utility from being employed to be positive, whereas a naive agent (wrongly) perceives it to be zero. Thus, a sophisticated agent is willing to accept a lower wage than a naive agent. The wage of a time-consistent agent is even lower, which, however, is solely driven by more first-period effort and the resulting higher rent that can be extracted.

Next, we show that the naive agent's higher wage also translates into a higher utility, compared to a sophisticated and a time-consistent agent. There, we follow the literature (see O'Donoghue and Rabin, 2001; DellaVigna and Malmendier, 2004; Gottlieb and Zhang, 2021) and compare *long-run realized* utility levels. We take the perspective of the period before the game starts, a "period 0", and denote the respective utilities by \hat{U}_0 .

Proposition 2 *A present-biased agent has a strictly larger long-run utility than a time-consistent agent, and a naive agent has a larger long-run utility than a sophisticated agent, i.e., $\hat{U}_0^{TC} = 0 < \hat{U}_0^S < \hat{U}_0^N$.*

¹⁴Note, however, that this does not imply that the total expected compensation of the time-consistent agent needs to be lower because it also involves a higher likelihood of receiving the benefit b .

The proof can be found in Appendix A.5.

For a time-consistent agent, \hat{U}_0^{TC} coincides with his period-1 utility, thus equals zero. The long-run utilities of a (naive or sophisticated) time-inconsistent agent, however, are strictly positive. This is because, from the perspective of “period 0”, there is no discounting between periods 1 and 2, whereas period-2 payoffs are discounted with β from the perspective of period 1. Thus, a time-inconsistent agent enjoys an extra “long-term” rent from his period-1 effort, which the principal cannot extract. Moreover, as discussed above, this extra rent also materializes from period-2 effort if the perspective of period 1 (or 0) is taken. Whereas the principal can extract this rent from a sophisticated agent, this is not possible with a naive agent, whose long-run utility therefore is even higher.

While the difference between time-consistent and time-inconsistent agent is the result of full bargaining power by the principal, the key reason for the difference between naive and sophisticated agent lies in the additional commitment assumption. The literature usually assumes that the principal has the power to commit to long-term contracts. Then, naive agents who misperceive their future preferences can “pay” to change their action once the future materializes, from the action that deemed optimal in the past to the action that is optimal now. In our setting, the misperception of his future preferences allows the naive agent to keep the “extra rent” generated by his time inconsistency. To fully grasp the role of commitment, assume that the second-period wage is given and the same for all agents. Then, the naive agent would not need to form expectations about future wages, but only about future effort. Since he is over-optimistic about his period-2 effort, he would accept a lower first-period wage than the sophisticated agent (see Section 6.1 for formal results).

5 Applications

We have shown that naiveté may actually protect an agent from excessive exploitation if the principal is not able to commit to long-term contracts. Having established the basic mechanism behind this result, we now elaborate on two examples where it might play a role. First, we assume that effort captures the agent’s on-the-job search activities to find a better-paid occupation. Second, we regard b as a performance-based bonus offered by the principal who benefits from the agent’s non-observable effort. We discuss both applications in turn and use them to derive further implications.

5.1 Effort as On-the-Job Search

Here, we take into account that the agent can look for a better job while being employed, with effort reflecting his search activities. More precisely, we assume that effort is identical to the probability with which the agent receives an outside job offer. This offer guarantees the agent a minimum payoff of b in all subsequent periods which thus constitutes his future outside option and reflects the agent's improved position on the labor market.¹⁵ b is drawn from an exogenous offer distribution $F(b)$ with support $[0, \bar{b}]$, $\bar{b} > 0$, similar to DellaVigna (2009).¹⁶ Consequently, if the agent has received an outside offer before, he will only benefit from future offers which exceed the previous realization of b . Therefore, wage and search effort in period 2 will also depend on the offer realization from period-1 search, which we denote by b_1 . Importantly, we do not take a stand on whether the agent actually moves to another employer or is retained by the principal. It only matters that an offer serves as a signal to the labor market that the agent's value is higher than previously assessed, and that he can keep on searching even after having received an offer.

Before formally analyzing this case, note that it relies on on-the-job search being more effective than search out of unemployment. Indeed, a number of reasons have been identified for why this is the case: A social stigma effect (see Biewen and Steffes, 2010, for evidence), a missing network (see Cingano and Rosolia, 2012, for evidence), the decay of human capital (Pissarides, 1992), or a higher likelihood of job termination by unemployed individuals (Nagypál, 2005) may reduce their chances of receiving a job offer. Generally, Mueller (2010) provides evidence that job search is more effective when being employed. Thus, on-the-job search can be viewed as a non-pecuniary benefit of being employed by the principal. For simplicity, we thus assume that search out of unemployment is not feasible, or alternatively, that its benefit is zero, in particular after rejecting the principal's offer.

Now, the agent can search in the first and second period, and we again proceed by backwards induction. In the second period, the time-inconsistent agent's optimal search effort, conditional on having received an offer $b_1 \geq 0$ in the first period, is given by:

$$e_2^k(b_1) = \beta \left(\int_{b_1}^{\bar{b}} b dF(b) - b_1(1 - F(b_1)) \right) \quad (1)$$

$$= \beta \int_{b_1}^{\bar{b}} (b - b_1) dF(b), \quad (2)$$

¹⁵If b instead constituted a one-time rent, the analysis would be equivalent to before.

¹⁶To always guarantee an interior solution, we assume $\mathbb{E}(b) < 1$.

for $k \in \{S, N\}$. Again, a time-consistent agent would search more and therefore, for a given $b_1 \geq 0$,

$$e_2^{TC}(b_1) > e_2^k(b_1) \quad \text{and} \quad w_2^{TC}(b_1) < w_2^k(b_1)$$

In contrast to the sophisticated agent, the naive agent perceives his period-2 preferences to be time consistent, thus $\tilde{e}_2^N(b_1) = e_2^{TC}(b_1)$ and $\tilde{w}_2^N(b_1) = w_2^{TC}(b_1)$. For all types of agents, however, $w_2(b_1) < 0$, and thus the ability to conduct on-the-job search allows the principal to reduce the period-2 wage below the agent's reservation utility (as previously derived by Board and Meyer-Ter-Vehn, 2015).

As before, in period 1 the naive agent's misperception of his future time preferences prevents the principal from extracting the "extra rent" stemming from his future time inconsistency. In contrast to the baseline model, however, this does not only affect the period-1 wage, but also period-1 effort which is determined by the utility difference between receiving and not receiving an outside offer. There, the "extra utility" of period-2 search from the perspective of $t = 1$, which is only taken into account by the sophisticated agent, again is important. This extra utility increases in future search, which itself is smaller after a first-period success, since then the agent only benefits from successful search if the resulting offer is larger than b_1 . Thus, the extra rent is larger if today's search turns out not to be successful, which reduces the sophisticated agent's motivation to exert search effort compared to the naive and the time-consistent agent.

Moreover, a potential search benefit mechanically decreases over time since it can be consumed for fewer remaining periods. Consequently, first-period search is, ceteris paribus, higher than second-period search. Thus, the naive and the time-consistent agent search more and are paid a lower wage in the first compared to the second period. This relationship is less clear for the sophisticated agent, though, because of the additional rent extraction by the principal.

This leads to the following implications for agents' search effort and long-run utilities:

Proposition 3 *Consider the on-the-job search application. Then in the first period,*

- *the sophisticated agent exerts less effort than the naive agent, who in turn exerts less than the time-consistent agent, i.e., $e_1^S < e_1^N < e_1^{TC}$.*
- *the naive agent's utility exceeds the sophisticated agent's utility, which in turn is larger than the time-consistent agent's utility of zero, i.e., $\hat{U}_0^{TC} = 0 < \hat{U}_0^S < \hat{U}_0^N$.*

The proof can be found in Appendix A.6.1.

Consequently, the insights concerning long-run utility from the baseline model also extend to this application: While time-inconsistency creates rents for the agent when the principal cannot commit to a long-term contract, it is only the naive agent who can reap those rents to the fullest. Moreover, whereas time inconsistency generally reduces search, it is the sophisticated agent who searches the least and therefore forgoes higher payoffs from better employment. The naive agent searches more *and* is better off.

5.2 b as Performance-Based Bonus

Here, we assume that the agent's effort benefits the principal but is not observable. Moreover, effort is identical to the probability with which a positive outcome is realized that has a value θ for the principal, with $\theta \in (0, 1)$. This event is verifiable, thus b corresponds to a performance-based bonus. The previous analysis can be applied to this case for a given bonus. Now, we also take into account that the principal is free to set b , and how a present bias and naiveté affects the principal's profits. Formally, the employment offer consists of $(w_t, b_t) \in \mathbb{R}^2$, where w_t is an upfront wage and b_t the bonus that can be obtained by period- t effort. There, note that the outcome realization happens at the beginning of period $t + 1$, which implies that also the bonus is effectively paid out one period after effort has been exerted. We argue that this does not contradict our no-commitment assumption because the performance contract is based on a verifiable measure and simply executed in the subsequent period. Alternatively, we could divide a period into two sub-stages, in which the first stage contains the wage payment and the agent's effort choice, whereas the second stage contains the realization of the output and payment of the bonus. Then, if sufficient time passes between the two stages, the bonus is discounted with β from the perspective of stage 1, and our results are the same as in the present setting. Finally, if the agent rejects the offer in any period, he still consumes his outside option utility of 0 from then on.

We again solve the game by backwards induction. In period 2, the principal's maximization problem with the naive and the sophisticated present-biased agent is given by

$$\begin{aligned} \max_{w_2, b_2} \quad & \pi + e_2(\theta - b_2) - w_2 \\ \text{s.t.} \quad & w_2 - \frac{1}{2}e_2^2 + \beta e_2 b_2 \geq 0, \end{aligned}$$

taking into account that effort is determined by

$$e_2 = \beta b_2.$$

w_2 is chosen to keep the agent to his outside option of zero, which yields a profit-maximizing bonus $b_2^k = \frac{\theta}{2-\beta}$, $k \in \{S, N\}$, for a time-inconsistent agent. In contrast, the optimal contract for a time-consistent agent involves a higher bonus $b_2^{TC} = \theta$. Moreover,

$$e_2^{TC} > e_2^k,$$

and

$$w_2^{TC} < w_2^k.$$

Note that, although the time consistent agent's upfront wage is smaller, his total expected compensation $w_2 + e_2 b_2$ exceeds the level of a time-inconsistent agent.

From the perspective of the first period, a time-inconsistent agent again enjoys an “extra rent” from period 2 effort. The principal can only extract this rent from the sophisticated agent, the naive misperceives his second-period preferences, consequently also his offered bonus ($\tilde{b}_2^N = b_2^{TC}$), effort ($\tilde{e}_2^N = e_2^{TC}$), and upfront wage ($\tilde{w}_2^N = w_2^{TC} < w_2^S$). An agent's perception of his second-period rent has no effect on his effort, though, thus also not on the optimal first-period bonus. Therefore, $e_1^S = e_1^N = \beta b$ and $b_1^S = b_1^N = \theta/(2-\beta)$, and the respective values are identical to their second-period counterparts. Nevertheless, whereas the naive agent expects to receive a period-2 utility of zero and thus does not accept a wage below $w_1^N = w_2^N$, the sophisticated agent suffers from an additional wage reduction, and $w_1^S < w_2^S$.

This yields the following implications regarding agents' long-run utilities and the principal's profits, where we denote the principal's present value of profits in period 1 by Π .

Proposition 4 *Assume that b is a performance-based bonus and can be set by the principal. Then,*

- *the naive agent's utility exceeds the sophisticated agent's utility, which in turn is larger than the time-consistent agent's utility of zero, i.e., $\hat{U}_0^{TC} = 0 < \hat{U}_0^S < \hat{U}_0^N$.*
- *the principal's profits are highest with a time-consistent agent; profits with a sophisticated agent exceed profits with a naive agent, i.e., $\Pi^{TC} > \Pi^S > \Pi^N > 0$.*

The proof can be found in Appendix A.6.2.

If the principal can set a bonus b , the same relationship of effort, wages, and long-run realized utilities between the different types of agents holds as with an exogenously given benefit.

The principal's profits are highest with a time-consistent agent. This is because he does not discount the bonus, thus it is cheaper to motivate him all else equal.

Consequently, incentives are stronger and the generated surplus is higher, which is completely extracted by the principal. If the agent is time-inconsistent, the principal's profits are higher with a sophisticated agent who receives a lower first-period wage but whose contract otherwise is identical to the contract of a naive agent.

6 Extensions

In this section we shed more light on key assumptions and explore the consequences of having alternative set-ups. There, we focus on the fact that our main results are driven by the naive agent's misperception of the future. Thus, all of the following extensions affect the extent of this misperception, with a reduction generally benefiting the principal but harming the naive agent. First, we allow the principal to (partially) commit to a future contract, which we argue also captures the consequences of (more or less stringent) employment protection. Second, we show that a moderate minimum wage may reduce the naive agent's misperception and consequently increase the rent that can be extracted from him. Third, we allow the agent to have private information regarding his naiveté. This might reduce the naive agent's misperception about his future wage and consequently harm him. In addition, it may have negative spillover effects on the sophisticated agent.

6.1 (Partial) Commitment and Employment Protection

In this section, we allow the principal to (partially) commit to future wages. To this end, we assume that the principal is able to fire the agent at the beginning of period 2, after the agent has potentially consumed the benefit b , however at firing cost k . The agent is still free to leave at any time. A higher k might reflect more stringent employment protection, or generally labor market regulation that protects workers. As long as b is exogenous and the principal only cares about wage payments, k is immaterial for the sophisticated and time-consistent agent. A higher k harms the naive agent, however, because it allows the principal to credibly commit to a second period wage that exceeds the level that maximizes profits from the perspective of the agent's first-period self. Recall that this wage equals $-b^2/2$, whereas the actually paid wage amounts to $-(\beta b)^2/2$. Now, assume that the principal promises a second period wage $\tilde{w}_2^N > -b^2/2$. Then, the agent takes into account that the principal could fire him at cost k and afterwards make a new wage offer $-b^2/2$.¹⁷ Paying a

¹⁷Alternatively, the agent could expect the principal to "renegotiate" the contract and pay k to the agent in exchange for accepting the respective wage cut

higher wage than promised of course is always possible, irrespective of the size of k .

Thus, an offer \tilde{w}_2^N made in period 1 is only credible (from the perspective of the agent's first-period self) if it satisfies

$$\tilde{w}_2^N \leq k + \left(-\frac{1}{2}b^2\right),$$

and the principal can credibly commit to the actual wage $-(\beta b)^2/2$ if

$$k \geq \frac{b^2(1-\beta^2)}{2}.$$

Otherwise $\tilde{w}_2^N = k - b^2/2$. This yields the following implications for the naive agent's first-period wage and his long-run utility.

Proposition 5 *Assume that the principal can commit to future contracts but deviate at cost k . Then, in the first period she offers a second-period wage*

$$\tilde{w}_2^N = \min \left\{ -\frac{1}{2}(\beta b)^2, k - \frac{1}{2}b^2 \right\},$$

which is increased to $w_2^N = -(\beta b)^2/2$ at the beginning of the second period. Moreover, the naive agent's first-period wage and his long-run utility

- *(weakly) decrease in k ,*
- *are smaller than the values of the sophisticated agent if k is sufficiently large,*
- *are always larger than the values of the time-consistent agent.*

The proof can be found in Appendix A.7.1.

The naive agent is harmed by higher termination costs k because the extent to which he underestimates the second-period wage is reduced. Thus, he is willing to accept a larger wage reduction in the first period. If k is sufficiently large, he is even worse off than the sophisticated agent. This is because the naive agent also overestimates his second-period effort and thus the rent he is then going to capture. The time-consistent agent anticipates the correct second-period wage anyway and thus is not affected by firing costs. Still, his long-run utility is below the levels of time-inconsistent agents because he does not enjoy the extra utility from a future time inconsistency.

Additional forces are at play if the principal can influence the agent's effort by endogenously setting the bonus. Then, commitment also affects the principal's relationship with the sophisticated agent. In fact, both the sophisticated agent

and the principal benefit from a higher k . The reason is that a higher bonus and consequently higher effort than the optimum from the perspective of the second period increase the agent's "extra utility" caused by his future time inconsistency. The principal can extract this extra utility in period 1, whereas the agent's long run self benefits once evaluating outcomes at the onset of the game. The result of Proposition 5 that the naive agent is harmed by commitment continues to hold with an endogenous bonus. A formal characterization of the consequences of (partial) commitment on outcomes with an endogenous bonus can be found in Section B.1 of Appendix B.

6.2 Minimum Wage

In this section, we show that our main mechanism can generate new insights on the effects of a minimum wage. Whereas a sufficiently high minimum wage in our setup benefits all agents, an intermediate level can actually harm a naive agent: Assume there is a minimum wage that exceeds the period-2 wage a naive agent expects to be paid but is below the wages he is paid in the first (and second) period. Then, a naive agent wrongfully anticipates a rent in the second period and consequently accepts a lower wage in the first. A sophisticated agent, on the other hand, would at all levels (weakly) benefit from a minimum wage. These results are collected in the following proposition.

Proposition 6 *Assume there is a minimum wage $\bar{w} \geq -b^2/2$. Then, a higher minimum wage has the following effects on wages and payoffs:*

- *For the naive agent, there exists a $\bar{w}^* \leq -(\beta b)^2/2$ such that $dw_1^N/d\bar{w} < 0$ and $d\hat{U}_0^N/d\bar{w} < 0$ if $\bar{w} \in [\hat{w}_2^N, \bar{w}^*)$. If $\bar{w} \geq \bar{w}^*$, $d\hat{U}_0^N/d\bar{w} > 0$.*
- *For the sophisticated and time-consistent agent, $d\hat{U}_0^k/d\bar{w} \geq 0$ and $dw_1^k/d\bar{w}$, $dw_2^k/d\bar{w} \geq 0$ for $k \in \{S, TC\}$.*

The proof can be found in Appendix A.7.2.

The naive agent can be harmed by a higher minimum wage because he underestimates his period-2 wage. Thus, it is possible that he expects the minimum wage to bind in the second period, although it actually is below his (first- and second-period) wage. Then, he is willing to accept a lower wage in the first period and experiences an effective utility reduction. The sophisticated agent can only benefit from a minimum wage; it increases his wages and limits the principal's ability to extract second-period rents. The same is true for the time-consistent agent.

The results for a naive agent indicate that a non-binding minimum wage might have *negative* spillover effects on higher wages. Previous literature has mostly focused on explanations for observed positive spillover effects, which include forces outside our model, such as firms wanting to preserve their wage distribution. However, there is evidence that spillover effects can indeed be negative. For example, Stewart (2012) examines the consequences of the British minimum wage. He finds that the growth of wages slightly above the minimum wage generally is *smaller* than what would have been expected without a minimum wage. Neumark et al. (2004) observe that, although immediate spillover effects are positive, lagged effects are strongly negative. Finally, Hirsch et al. (2015) provide indicative evidence that wages above the minimum wage increase less strongly than they would have without the minimum wage.

To conclude this section, note that a minimum wage can also have consequences on our applications, in particular on effort as on-the-job search. Then, a minimum wage is more likely to bind after a failure than after a success. Therefore, we would expect a minimum wage to reduce search effort – and consequently turnover – because the benefits of an unaltered employment relationship go up. This holds in particular for the naive agent who might wrongly perceive a minimum wage to bind in the second period.

Less search on the labor market likely goes hand in hand with a reduction in labor *turnover*. Whereas there is abundant evidence on negative turnover effects, a number of theoretical explanations can already yield explanations. These explanations only hold for *binding* minimum wages, though, whereas we would also predict negative turnover effects of a *non-binding* minimum wage, which the naive agent wrongly perceives to bind in the future. To the best of our knowledge, this aspect has so far not been explored empirically. Some indicative evidence is provided by Hirsch et al., 2015, who find that the negative effects of a minimum wage on turnover are not necessarily increasing in compliance costs, which are a measure of the extent to which minimum wages bind.

6.3 Asymmetric Information

Our results rely on the principal knowing the agent’s the agent’s extent of naiveté and, closely related, on the naive agent’s belief that the principal shares his own perception about his type. We now refer to those baseline model assumptions as *symmetric* information.¹⁸

¹⁸Note that our previous analysis does not assume symmetric information in the strict sense since the naive agent does not share the principal’s belief about his own future preferences. Rather, it represents a form of non-common priors, as stated by Eliaz and Spiegler (2006).

In this section, we consider the case of asymmetric information in the sense that the principal is not able to observe the agent’s naiveté, neither at the time of contracting nor at any later point in time. Hence, as in Eliaz and Spiegler (2006), we focus on time-inconsistent agents and consider the level of β as common knowledge, but allow for uncertainty about the agent’s extent of naiveté. We assume that the principal is randomly matched with an agent before the employment relationship starts, and that the agent is naive with probability α and sophisticated with probability $1 - \alpha$, where α is known to the principal. Moreover, if the principal abstains from making an offer to the agent, or if the agent does not accept her offer, she cannot employ him in later periods.

In our setting without long-term commitment, asymmetric information about the agent’s naiveté implies that, in order to anticipate future wages, the agent not only has to form beliefs about his own future present bias, but also about the (principal’s assessment of the) distribution of other agents’ present bias. Such beliefs do not have to be specified in principal-agent models with full commitment where the principal can guarantee a stream of contingent future payments which only depend on the agent’s own actions and thus his own present bias (see Eliaz and Spiegler, 2006; Englmaier et al., 2016; Gottlieb and Zhang, 2021). In the following, we thus build upon games of present-biased players where such beliefs have been formulated. There, it is generally assumed that the sophisticated agent – just like the principal – knows that all agents are time-inconsistent and is aware of the share of naive and sophisticated agents in the population.

Regarding the naive agent’s inter-player expectations we consider two alternatives, while we always stick to the assumption that the naive agent believes the principal to share his own perception:

Assumption 1a (Interpersonal Naiveté) *The naive agent believes that all other agents are time-consistent in the future.*

Assumption 1b (Interpersonal Sophistication) *The naive agent believes that all other agents are time-inconsistent in the future.*

Assumption 1a follows Sarafidis (2006), whose analysis of a bargaining game between time-inconsistent players is based on such *interpersonal naiveté*. The presumption that a naive agent believes all other agents to be time consistent in the future as well reflects a situation in which a naive agent does not perceive himself to be distinct. The contrary, *interpersonal sophistication*, is the foundation of Assumption 1b and follows Akin (2007). It is consistent with evidence that agents are over-confident (see Santos-Pinto and de la Rosa, 2020) and perceive themselves to be less present biased than others (see Fedyk, 2018).

Equipped with these assumptions, we use the solution concept of *Naive Backwards Induction* as propagated by Sarafidis (2006). There, naive agents' optimal strategies are based on their perceptions of other agents' preferences and thus actions.¹⁹ In the following, we characterize potential outcomes, how they rely on the share of naive agents, α , and on the naive agent's beliefs about the present bias of other agents as described in Assumptions 1a and 1b.

We assume that the principal's base profits when employing an agent are π (for example generated by actions that can be contractually specified), and that her profits when not employing an agent are $\underline{\pi} < \pi$. Besides, she only cares about the wage payments. Now, the symmetric-info result that realized outcomes in period 2, as well as the optimal effort level in period 1, solely rely on β and not the agent's naiveté also extends to the present setting with asymmetric information. Only the naive agent's period-1 belief regarding period-2 outcomes depends on the above assumptions. This belief ultimately affects the naive agent's reservation wage in period 1 and thereby the principal's profits.

Note that a separating contract in which different types of agents accept different wages in the first period does not exist because the principal cannot commit to condition her second-period wage offer on any first-period decision. Hence, we only consider pooling equilibria, where we impose as tie-breaking rule that, whenever the principal is indifferent between employing both types of agents and employing only one type of agent, she chooses the former. We start with a characterization of the results under *interpersonal naiveté*, that is, we impose Assumption 1a.

Proposition 7a *Consider the case with asymmetric information and suppose Assumption 1a holds. Then the principal's expected profits are smaller than with symmetric information. Moreover, there exists a threshold $\bar{\alpha} \in (0, 1)$ such that:*

- *If $\alpha \geq \bar{\alpha}$, the principal employs both types of agents.*
- *If $\alpha < \bar{\alpha}$, the principal employs only the sophisticated agent.*

The proof can be found in Appendix A.7.3.

Under Assumption 1a the naive agent also perceives all other agents to be time consistent in the future. Therefore, he believes that the principal will offer the symmetric information wage \tilde{w}_2^N in period 2, which is lower than the actually offered wage w_2^S . In turn, the naive agent's reservation wage in period 1 will be the same as in the symmetric information case. Thus, the principal has to decide

¹⁹More precisely, applying naive backwards induction to games of time-inconsistent and potentially naive agents usually also requires the formulation of higher order beliefs. This is not necessary in our setting, thus we refrain from doing so.

whether to offer the reservation wage of the naive agent w_1^N which is also accepted by the sophisticated agent, or the lower reservation wage of the sophisticated agent w_1^S . She is only willing to do the former if the share of naive agents is sufficiently large. In any case, the principal is worse off than in the symmetric information case. Either, she has to grant the sophisticated agent a wage which exceeds his reservation wage, or she risks to not fill the vacancy if the agent turns out to be naive.

The situation is different if we impose *interpersonal sophistication*, that is, Assumption 1b.

Proposition 7b *Consider the case with asymmetric information and suppose Assumption 1b holds. Then the principal's profits are larger than with symmetric information. Moreover, there exists a threshold $\bar{\alpha} \in (0, 1)$ such that:*

- *If $\alpha > \bar{\alpha}$, the principal employs only the naive agent.*
- *If $\alpha \leq \bar{\alpha}$, the principal's employs both types of agents .*

The proof can be found in Appendix A.7.3.

Under Assumption 1b, the naive agent perceives all other agents to be time-inconsistent in period 2. Therefore, he expects the principal to offer the wage w_2^S , which exceeds his own anticipated reservation wage, \tilde{w}_2^N . Thus, the naive agent has overoptimistic beliefs about his future payoff. This implies that his reservation wage in period 1 is lower than under symmetric information, and it is lower than the sophisticated agent's first-period wage, $\hat{w}_1^N < w_1^S$. Thus, the principal has to choose whether to offer the reservation wage of the naive agent, which will be rejected by the sophisticated agent, or the higher reservation wage of the sophisticated agent, which will also be accepted by the naive agent. She will do the former if the share of naifs in the population is sufficiently large.

In contrast to the result under Assumption 1a, the principal will always be *better off* than under symmetric information. The reason is that now, the naive agent is also willing to work for the sophisticated agent's reservation wage, which is the same as under symmetric information and smaller than the naive agent's symmetric-info reservation wage. If she decides to exclude the sophisticated agent, then only because her profits would be even larger due to a low \hat{w}_2^N and a high likelihood of facing a naive agent. In this case, naive agents have a negative spillover effect on sophisticated agents because the principal refrains from employing them.

To conclude, the principal may actually benefit from being ignorant about the agent's naiveté, or more precisely, from not being able to discriminate between different agents. The result also provides interesting insights about how learning about his or others' present bias affects the naive agent. Whereas, under symmetric

information, the principal would clearly benefit from inducing him to learn about his present bias, the case is different with asymmetric information. Then, she would prefer the naive agent to correctly assess *other agents'* present bias and keep him naive in order to maintain his overoptimistic regarding future effort.

7 Robustness

We now turn to a discussion of the robustness of our results once we relax some assumptions not discussed before. First, we show that our results do not only hold for the extreme cases of full naiveté and sophistication, but are continuous in the extent of naiveté. Second, we demonstrate that our mechanism is not specific to three periods, but extends to larger time horizons. Third, allowing for additional discounting and present bias between agreeing to a contract and effort exertion also does not change our main insight.

7.1 Partially Naive Agent

In the main part of this paper we have focused on two extreme cases: Full naiveté or full sophistication. However, workers may anticipate their future present bias to some extent, but not fully. In this section we consider the case where an agent may exhibit partial naiveté and show how he is affected by a marginal increase in naiveté. Hence, while as before, utilities in the next period are discounted with β , the agent expects to discount the future with $\tilde{\beta} \in (\beta, 1)$ from the next period onward. Sections 4.1 and 4.2 have therefore covered the cases $\tilde{\beta} = \beta$ and $\tilde{\beta} = 1$.

All realized outcomes in the second period, as well as effort in the first, are independent of $\tilde{\beta}$, hence are equivalent to the respective values with a sophisticated or fully naive agent. From the perspective of period 1, a partially naive agent expects to maximize $-(e_2)^2/2 + \tilde{\beta}e_2b$ in period $t = 2$. This implies that the partially naive agent expects to choose an effort level \tilde{e}_2^{PN} which is characterized by

$$\tilde{e}_2^{PN} = \tilde{\beta}b.$$

Furthermore, in period 1 the partially naive agent anticipates a second-period wage offer

$$\tilde{w}_2^{PN} = -\frac{1}{2}(\tilde{\beta}b)^2.$$

As before, the agent's first-period reservation wage is determined by his perceptions of future outcomes, not their true realizations. The following Lemma focuses on

$\beta \geq \frac{1}{2}$ which is in line with empirical estimates of present bias which generally find values of β that are not too small (see Imai et al. (2020))

Lemma 4 *Assume the agent is partially naive with $\tilde{\beta} \in (\beta, 1)$ and $\beta \geq \frac{1}{2}$. Then,*

- *the period-1 wage that is lower than the period-2 wage, i.e. $w_1^{PN} = -\frac{1}{2}(\beta b)^2 - \beta \tilde{\beta}(1 - \tilde{\beta})b^2 < w_2^{PN}$.*
- *A higher extent of naiveté increases the period-1 wage, i.e. $\frac{\partial w_1^{PN}}{\partial \tilde{\beta}} > 0 \forall \tilde{\beta} \in (\beta, 1)$. Moreover, $\lim_{\tilde{\beta} \rightarrow \beta} w_1^{PN} = w_1^S$ and $\lim_{\tilde{\beta} \rightarrow 1} w_1^{PN} = w_1^N$.*

The proof can be found in Appendix A.8.1.

The relationship between period-1 and period-2 wages is the same as for the sophisticated agent. Because the difference between the partially naive agent's current preference over future effort and the perceived future effort is monotonously and continuously decreasing in the extent of naiveté, so is his expected wage offer and therefore his current reservation wage.²⁰

Finally, when comparing a partially naive agent to a time-consistent agent in terms of long-run realized utility, a similar relationship emerges. Corollary 1 directly follows from combining Lemma 4 and Proposition 2, again focusing on the arguably more realistic case $\beta \geq \frac{1}{2}$.

Corollary 1 *Assume the agent is partially naive with $\tilde{\beta} \in (\beta, 1)$ and $\beta \geq \frac{1}{2}$. $\hat{U}_0^{TC} = 0 < \hat{U}_0^S < \hat{U}_0^{PN} < \hat{U}_0^N$. Moreover, \hat{U}_0^{PN} is strictly increasing in $\tilde{\beta}$.*

A larger extent of naiveté makes a present-biased agent better off. This result differs from the behavioral IO literature. There, a discontinuity at $\tilde{\beta} = \beta$ is often observed. If firms have commitment power, (partially) naive agents have to "pay" for a reversion of their actions, and for their payoffs it only matters that a misperception of actions happens, not its extent.

7.2 Discounting Between Each Step

When the agent assesses a period's wage and effort at the beginning of a period, both relate to the "present". However, evidence indicates that a present bias is only about very close events (O'Donoghue and Rabin, 2015). For example, Augenblick (2018) shows that the β discounts consumption already a few hours away and in particular that consumption more than a few days away is not included in the "present" an individual is biased towards. We now discuss the robustness of our results if we

²⁰Note that, allowing for $\beta < \frac{1}{2}$ would only affect the comparative static result in Lemma 4: The wage of the partially naive agent decreases in $\tilde{\beta}$ for $\tilde{\beta} \in (\beta, \frac{1}{2})$, but then increases for $\tilde{\beta} \in [\frac{1}{2}, 1)$.

incorporate these aspects into our model. To do so, we split each period into two steps and discount period- t effort costs with β already *at the beginning* of period t , after receiving w_t . For example, at the beginning of period 2 an agent's utility is

$$w_2 + \beta \left[-\frac{1}{2}e_2^2 + e_2b \right],$$

whereas at the time when the effort choice is made his utility amounts to

$$-\frac{1}{2}e_2^2 + \beta e_2b.$$

Then, the naive agent also overestimates his effort in $t = 2$ when assessing the period's wage offer. He expects to exert effort $\tilde{e}_2^N = b$ and accepts a wage satisfying $w_2^N + \beta [-\tilde{e}_2^2/2 + \tilde{e}_2b] = 0$. Thus, the naive agent's *realized* period-2 utility now is negative from the perspective of period 2 (whereas the sophisticated agent still only accepts a wage that yields a non-negative utility). It continues to be positive from the perspective of earlier periods, though. In total, discounting future payoffs after each step reduces the naive agent's utility compared to our main model because he overestimates his effort also at the beginning of a period and thus accepts a lower wage. Still, it is possible that his long-run utility exceeds the level of the sophisticated agent, which is captured in the following proposition.

Proposition 7 *Assume that time-inconsistent agent discounts future payoffs also after receiving w_t and before exerting effort. Then, the long-run utilities of both, sophisticated and naive agent, are positive. The naive agent's long-run utility exceeds the sophisticated agent's long-run utility if β is sufficiently large.*

The proof can be found in Appendix A.8.2.

7.3 Infinite Time Horizon

In this section, we show that our results do not rely on a particular number of periods but also hold within an infinite time horizon. Our setup is as before, only that we introduce “standard” exponential discounting between periods, captured by the discount factor δ .²¹

Then, an agent's utility in a period t equals

$$U_t = w_t - \frac{e_t^2}{2} + \beta\delta \left(e_t b + \hat{U}_{t+1} \right),$$

²¹without such additional discounting, future payoffs would not be bounded.

with

$$\hat{U}_{t+1} = \sum_{\tau=t+1}^{\infty} \delta^{\tau-(t+1)} \left(w_{\tau} - \frac{e_{\tau}^2}{2} + \delta e_{\tau} b \right).$$

As before, effort is only determined by the chance to obtain next period's bonus, therefore

$$e = \beta \delta b$$

in every period t . The naive agent, however, expects to exert effort

$$\tilde{e}^N = \delta b$$

in any future period.

For the following, note that stationary outcomes in which wages are the same in every period are optimal with an exogenous benefit b .²² Thus we can omit time subscripts.

Now, the *naive* agent's misperception of his future time preferences induces him to believe that his future wage \tilde{w} satisfies

$$\tilde{U}^N = \frac{\tilde{w}^N - \frac{(\tilde{e}^N)^2}{2} + \delta \tilde{e}^N b}{1 - \delta} = 0,$$

and

$$\tilde{w}^N = -\frac{(\delta b)^2}{2}.$$

Therefore, in any period he anticipates a continuation utility of zero and does not accept a wage below

$$w^N = -\frac{(\beta \delta b)^2}{2},$$

which yields a long-run utility

$$\hat{U}^N = \frac{\beta (\delta b)^2 (1 - \beta)}{1 - \delta}.$$

The *sophisticated* agent correctly anticipates future outcomes. Therefore, his wage equals

$$w^S = -\frac{(\beta \delta b)^2}{2} \frac{1 + \delta (1 - \beta)}{1 - \delta (1 - \beta)},$$

which yields a long-run utility

²²Realized and perceived future wages still differ for the naive agent, stationarity then means that all perceived future wages are the same, as well as all realized wages.

$$\hat{U}^S = \frac{\beta (\delta b)^2 (1 - \beta)}{1 - \delta (1 - \beta)} < \hat{U}^N.$$

All this implies that trade-offs do not change once the time horizon is infinite (or finite with any number of periods). The naive agent still underestimates his future wage, which makes it impossible for the principal to extract the “extra rent” stemming from his time inconsistency, and is better off in the long run. Finally, a *time-consistent* agent’s long-run utility is obtained by taking \hat{U}^S and setting $\beta = 1$, thus equals zero.

Regarding our applications, an infinite time horizon would not affect the trade-offs associated with on-the-job search. With an endogenous bonus and a sophisticated agent, however, the principal is able to *endogenously* commit to a future bonus that exceeds the myopic level. In Section 6.1, we have shown that this possibility exists with formal commitment, raising the “extra rent” of the agent’s future time inconsistency and thus increasing profits and the agent’s long-run utility. Without formal commitment but an infinite time horizon and a sufficiently high discount factor δ , such an outcome can be sustained in an equilibrium in which any deviation is punished by a reversion to the myopic optimum in all future periods.²³ A formal derivation of such an equilibrium can be found in Section B.2 of Appendix B.

8 Discussion and Conclusion

We have shown that present-biased agents can benefit from being naive if firms are not able to commit to long-term contracts. Commitment, for example provided by employment protection, harms the naive agent but can benefit the principal and the sophisticated agent. However, even if commitment is generally feasible, the insights developed in our model can still be relevant. Certainly there are also forces outside our model that limit the benefits of commitment, e.g., uncertainty regarding future prospects. For example, assume there is a chance that the principal’s profits from employing the agent might drop in any period (such as due to demand fluctuations or productivity shocks), and the principal would rather terminate the relationship in that case. Then, she would have to trade off the reduced flexibility of commitment with the possibility to exploit the naive agent. Moreover, even if we take into account that commitment to future wages often is possible, such commitment generally does not hold for arbitrarily long time periods. At some point, the principal has the

²³A similar mechanism is found in Fahn and Hakenes (2019), who show that teamwork allows time-inconsistent (and sophisticated) individuals to overcome their self-control problems.

opportunity to end the contract and then negotiate the terms of a continuation.

Finally, we want to discuss the role of the competitive environment and consequently principal's ability to make take-it-or-leave-it offers. The consequences of relaxing this assumption would depend on players' alternative opportunities in every period, as well as potential matching frictions. A general characterization is beyond the scope of this paper, some insights can nevertheless be derived in the following sketch. Assume there are many principals and agents, and each principal can hire exactly one agent. Moreover, there is random matching without frictions in every period. Being matched with an agent, the principal observes his type (sophisticated or naive) and offers a contract. If there are more agents than principals, the principal can make a take-it-or-leave-it offer but has to wait until the next period if she wants to be matched with a different agent; with more principals than agents, the agent can make a take-it-or-leave-it offer. In the former case, outcomes are as in our model. The reason is that principals will in every period offer contracts that are accepted by all types agents: even if matched with a naive agent, waiting for the next period to then potentially employ a sophisticated agent is not profitable because second-period profits are the same with both types. Thus, an agent who rejects a first-period offer remains unmatched in the second period which is one of the causes of the payoff differences between naive and sophisticated agent in our setting. If agents can make take-it-or-leave-it offers, though, the situation is different. Then, wages are equal to a principal's per-period net value of employing the agent. Because this net value is independent of the agent's effort as well as his naiveté, also wages are the same for both types. Consequently, both have the same long-run utilities, only their perceived utilities at the beginning of the first period are different. Thus, competition for agents lets the naive agent's advantage disappear. With matching frictions, we would conjecture that the outcome is a weighted average of both cases, with the payoff difference between naive and sophisticated agent decreasing in the number of principals.

Concluding, the role of commitment of firms who deal with present-biased individuals has so far only played a limited role in the literature. We hope that our paper can serve as one step towards a better understanding of the importance of commitment, in particular in labor markets.

References

- Akin, Zafer**, “Time inconsistency and learning in bargaining games,” *International Journal of Game Theory*, 2007, *36* (2), 275–299. 6.3
- Amador, Manuel, Iván Werning, and George-Marios Angeletos**, “Commitment vs. flexibility,” *Econometrica*, 2006, *74* (2), 365–396. 2
- Augenblick, Ned**, “Short-term time discounting of unpleasant tasks,” *Unpublished manuscript*, 2018. 7.2
- Biewen, Martin and Susanne Steffes**, “Unemployment persistence: Is there evidence for stigma effects?,” *Economics Letters*, 2010, *106*, 188–190. 5, 5.1
- Bjelland, Melissa, Bruce Fallick, John Haltiwanger, and Erika McEntarfer**, “Employer-to-Employer Flows in the United States: Estimates Using Linked Employer-Employee Data,” *Journal of Business & Economic Statistics*, 2011, *29* (4), 493–505. 2
- Board, Simon and Moritz Meyer-Ter-Vehn**, “Relational Contracts in Competitive Labour Markets,” *The Review of Economic Studies*, 2015, *82* (2), 490–534. 5.1
- Bond, Philip and Gustav Sigurdsson**, “Commitment contracts,” *The Review of Economic Studies*, 2018, *85* (1), 194–222. 2
- Bubb, Ryan and Patrick L. Warren**, “An Equilibrium Theory of Retirement Plan Design,” *American Economic Journal: Economic Policy*, May 2020, *12* (2), 22–45. 2
- Cingano, Federica and Alfonso Rosolia**, “People I Know: Job Search and Social Networks,” *Journal of Labor Economics*, 2012, *30* (2), 291–332. 5, 5.1
- DellaVigna, Stefano**, “Psychology and economics: Evidence from the field,” *Journal of Economic literature*, 2009, *47* (2), 315–72. 1, 5.1
- **and M. Daniele Paserman**, “Job search and impatience,” *Journal of Labor Economics*, 2005, *23* (3), 527–588. 2
- **and Ulrike Malmendier**, “Contract Design and Self-Control: Theory and Evidence*,” *The Quarterly Journal of Economics*, 2004, *119* (2), 353–402. 4.3
- Eliaz, Kfir and Ran Spiegler**, “Contracting with Diversely Naive Agents,” *The Review of Economic Studies*, 2006, *73* (3), 689. 3, 1, 2, 9, 18, 6.3

- Englmaier, Florian, Matthias Fahn, and Marco A. Schwarz**, “Long-Term Employment Relations when Agents are Present Biased,” 2016. Working Paper. 2, 3, 6.3
- Fahn, Matthias and Hendrik Hakenes**, “Teamwork as a Self-Disciplining Device,” *American Economic Journal: Microeconomics*, November 2019, 11 (4), 1–32. 23
- Fallick, Bruce C. and Charles A. Fleischman**, “The Importance of Employer-to-Employer Flows in the U.S. Labor Market,” 2001. Working Paper. 2
- Fedyk, Anastassia**, “Asymmetric naivete: Beliefs about self-control,” *Available at SSRN 2727499*, 2018. 1, 6.3
- Gautier, Pieter A., Coen N. Teulings, and Aico Van Vuuren**, “On-the-Job Search, Mismatch and Efficiency,” *Review of Economic Studies*, 2010, 77 (1), 245–272. 2
- Gilpatric, Scott M.**, “Present-biased preferences, self-awareness and shirking,” *Journal of Economic Behavior & Organization*, 2008, 67 (3), 735–754. 2, 9
- Gottlieb, Daniel and Xingtang Zhang**, “Long-Term Contracting With Time-Inconsistent Agents,” *Econometrica*, 2021, 89 (2), 793–824. 2, 4.3, 6.3
- Grubb, Michael D**, “Overconfident consumers in the marketplace,” *Journal of Economic Perspectives*, 2015, 29 (4), 9–36. 8
- Heidhues, Paul and Botond Köszegi**, “Exploiting Naivete about Self-Control in the Credit Market,” *American Economic Review*, 2010, 100 (5), 2279–2303. 7, 2
- **and** –, “Naïveté-Based Discrimination,” *The Quarterly Journal of Economics*, 11 2017, 132 (2), 1019–1054. 2
- **and** –, “Behavioral industrial organization,” *Handbook of Behavioral Economics: Applications and Foundations 1*, 2018, 1, 517–612. 1
- Hirsch, Barry T., Bruce E. Kaufman, and Tetyana Zelenska**, “Minimum Wage Channels of Adjustment,” *Industrial Relations: A Journal of Economy and Society*, 2015, 54 (2), 199–239. 1, 6.2
- Hoffman, Mitchell and Stephen V. Burks**, “Worker overconfidence: Field evidence and implications for employee turnover and firm profits,” *Quantitative Economics*, 2020, 11 (1), 315–348. 2

- Imai, Taisuke, Tom A Rutter, and Colin F Camerer**, “Meta-Analysis of Present-Bias Estimation using Convex Time Budgets,” *The Economic Journal*, 09 2020, *131* (636), 1788–1814. 7.1
- Kaur, Supreet, Michael Kremer, and Sendhil Mullainathan**, “Self-control and the development of work arrangements,” *American Economic Review*, 2010, *100* (2), 624–28. 1
- , – , and – , “Self-control at work,” *Journal of Political Economy*, 2015, *123* (6), 1227–1277. 1, 2
- Kőszegi, Botond**, “Behavioral Contract Theory,” *Journal of Economic Literature*, 2014, *52* (4), 1075–1118. 1, 7, 8
- Laibson, David**, “Golden Eggs and Hyperbolic Discounting*,” *The Quarterly Journal of Economics*, 1997, *112* (2), 443–478. 1, 3
- Mortensen, Dale T**, “Equilibrium Unemployment with Wage Posting: Burdett-Mortensen Meet Pissarides,” in “Panel Data and Structural Labour Market Models,” Emerald Group Publishing Limited, 2000. 2
- Moscarini, G.**, “Job Matching and the Wage Distribution,” *Econometrica*, 2005, *73*, 481–516. 2
- Mueller, Andreas**, “On-the-job search and wage dispersion: New evidence from time use data,” *Economics Letters*, 2010, *109* (2), 124–127. 5, 5.1
- Nagypál, Éva**, “Amplification of Labor Market Fluctuations: Why Vacancies Don’t Like to Hire the Unemployed?,” Technical Report, Society for Economic Dynamics 2005. 5.1
- , “Worker reallocation over the business cycle: The importance of employer-to-employer transitions,” *Manuscript, Northwestern Univ*, 2008. 2
- Neumark, David, Mark Schweitzer, and William Wachter**, “Minimum Wage Effects throughout the Wage Distribution,” *Journal of Human Resources*, 2004, *39* (2), 425–450. 1, 6.2
- O’Donoghue, Ted and Matthew Rabin**, “Doing It Now or Later,” *The American Economic Review*, 1999, *89* (1), 103–124. 2, 3, 3
- and – , “Incentives for Procrastinators,” *The Quarterly Journal of Economics*, 1999, *114* (3), 769–816. 1, 2

- and –, “Choice and Procrastination,” *The Quarterly Journal of Economics*, 2001, *116* (1), 121–160. 4.3
- and –, “Present Bias: Lessons Learned and to Be Learned,” *American Economic Review*, May 2015, *105* (5), 273–79. 7.2
- Paserman, M. Daniele**, “Job search and hyperbolic discounting: Structural estimation and policy evaluation,” *The Economic Journal*, 2008, *118* (531), 1418–1452. 2
- Pissarides, C.A.**, “Search Unemployment with On-The-Job Search,” *Review of Economic Studies*, 1994, *61*, 457–475. 2
- Pissarides, Christopher A.**, “Loss of skill during unemployment and the persistence of employment shocks,” *Quarterly Journal of Economics*, 1992, *107*, 1371–1391. 5.1
- Santos-Pinto, Luis and Leonidas Enrique de la Rosa**, “Overconfidence in labor markets,” *Handbook of Labor, Human Resources and Population Economics*, 2020, pp. 1–42. 6.3
- Sarafidis, Yianis**, “Games with time inconsistent players,” *Available at SSRN 954394*, 2006. 6.3
- Stewart, Mark B.**, “Wage inequality, minimum wage effects, and spillovers,” *Oxford Economic Papers*, 02 2012, *64* (4), 616–634. 1, 6.2
- Yılmaz, Murat**, “Repeated moral hazard with a time-inconsistent agent,” *Journal of Economic Behavior & Organization*, 2013, *95*, 70–89. 2

A Appendix

A.1 Proof of Lemma 1.

Conditional on having accepted the principal's employment offer at the beginning of $t = 1$, the agent chooses effort to maximize,

$$-\frac{1}{2}(e_1)^2 + \left[e_1 b + w_2^{TC} - \frac{1}{2}(e_2^{TC})^2 + e_2^{TC} b \right].$$

Since effort does not affect the expected wage in the second period, the problem boils down to maximizing $-\frac{1}{2}(e_1)^2 + e_1 b$, yielding $e_1^{TC} = b$, as well as $w_1^{TC} = -\frac{1}{2}b^2 < 0$, hence $e_1^{TC} = e_2^{TC}$ and $w_1^{TC} = w_2^{TC}$. ■

A.2 Proof of Lemma 2.

e_1^S maximizes

$$-\frac{1}{2}e_1^2 + \beta \left[e_1 b + w_2 - \frac{1}{2}e_2^2 + e_2 b \right]$$

hence

$$e_1^S = \beta b.$$

It follows that $e_1^S = e_2^S = \beta b$.

Plugging in $w_2^S = -\frac{1}{2}(\beta b)^2$ and $e_2^S = \beta b$, w_1^S is set to satisfy $U_1^S = w_1^S - \frac{1}{2}(e_1^S)^2 + \beta [e_1^S b + (1 - \beta) \beta b^2] = 0$, hence

$$w_1^S = \frac{1}{2}(e_1^S)^2 - \beta [e_1^S b + (1 - \beta) \beta b] < \frac{1}{2}(e_1^S)^2 - \beta e_1^S b = -\frac{1}{2}(\beta b)^2. \quad \blacksquare$$

A.3 Proof of Lemma 3

A fully naive agent perceives his first-period utility to be

$$\tilde{U}_1^N = w_1 - \frac{1}{2}(e_1)^2 + \beta \left[e_1 b + \tilde{w}_2^N - \frac{1}{2}(\tilde{e}_2^N)^2 + \tilde{e}_2^N b \right].$$

Making use of $\tilde{w}_2^N = \frac{1}{2}(\tilde{e}_2^N)^2 - \tilde{e}_2^N b$, this becomes

$$\tilde{U}_1^N = w_1 - \frac{1}{2}(e_1)^2 + \beta e_1 b.$$

The Lemma immediately follows. ■

A.4 Proof of Proposition 1.

As shown in Lemmas 2 and 3, period-1 efforts of a sophisticated agent and naive agent equal

$$e_1^S = \beta b = e_1^N.$$

As shown in Lemma 1, period-1 effort of a time-consistent agent equals

$$e_1^{TC} = b.$$

Hence, the first part of the Proposition immediately follows.

As to the second part, from Lemmas 2 and 3, it follows that

$$w_1^S = -\frac{1}{2}(\beta b)^2(3 - 2\beta) < -\frac{1}{2}(\beta b)^2 = w_1^N$$

Moreover, $w_1^{TC} < w_1^S$ since

$$\begin{aligned} -\frac{1}{2}b^2 &< -\frac{1}{2}(\beta b)^2(3 - 2\beta) \\ 1 &> \beta^2(3 - 2\beta) \end{aligned}$$

where the second line follows from multiplying the first by -2 and dividing it by b^2 . The inequality is true because the right hand side is strictly increasing in β and $\lim_{\beta \rightarrow 1} \beta^2(3 - 2\beta) = 1$. ■

A.5 Proof of Proposition 2.

Any agent's long-run realized utility is given by:

$$\hat{U}_0 = w_1 - \frac{1}{2}e_1^2 + \left[e_1 b + w_2 - \frac{1}{2}e_2^2 + e_2 b \right]$$

Evaluating the realized utility at the equilibrium contract and effort level, w_t^{TC}, e_t^{TC} , $t \in \{1, 2\}$, of the time-consistent agent, we get

$$\hat{U}_0^{TC} = -\frac{1}{2}b^2 - \frac{1}{2}b^2 + b^2 - \frac{1}{2}b^2 - \frac{1}{2}b^2 + b^2 = 0$$

Similarly, evaluating the realized utility at the equilibrium contract and effort

level, w_t^S, e_t^S , $t \in \{1, 2\}$, of the sophisticated agent, we get

$$\begin{aligned}\hat{U}_0^S &= -\frac{1}{2}(\beta b)^2(3 - 2\beta) - \frac{1}{2}(\beta b)^2 + \beta b^2 - \frac{1}{2}(\beta b)^2 - \frac{1}{2}(\beta b)^2 + \beta b^2 \\ &= -(\beta b)^2(2 - \beta) + \beta(2 - \beta)b^2 \\ &= (2 - \beta)\beta(1 - \beta)b^2 > 0.\end{aligned}$$

Finally, the realized utility at the equilibrium contract and effort level, w_t^N, e_t^N , $t \in \{1, 2\}$, of the naive agent, amounts to

$$\begin{aligned}\hat{U}_0^N &= -\frac{1}{2}(\beta b)^2 - \frac{1}{2}(\beta b)^2 + \beta b^2 - \frac{1}{2}(\beta b)^2 - \frac{1}{2}(\beta b)^2 + \beta b^2 \\ &= 2\beta(1 - \beta)b^2 > (2 - \beta)\beta(1 - \beta)b^2 > 0.\end{aligned}$$

The first inequality is intuitive as the only difference between the contract of the naive and the sophisticated agent is that the sophisticated receives a smaller wage in the first period. ■

A.6 Applications

A.6.1 Proof of Proposition 3 (Effort as On-the-job Search)

For the following analysis, we do not specifically analyze the principal; a successful counteroffer increases the agent's net utility from being employed by the same amount, no matter whether he stays with the principal or moves to another employer.

For the equilibrium outcomes in stage 2, we start with the analysis of the time-inconsistent agent. Conditional on having a per-period outside option b_1 , the sophisticated as well as naive agent's search level maximizes

$$-\frac{1}{2}e_2^2 + \beta \left[e_2 \left(\int_{b_1}^{\bar{b}} b dF(b) + b_1 F(b_1) \right) + (1 - e_2)b_1 \right].$$

Thus, the agent's optimal period-2 effort $e_2^k(b_1)$, $k \in \{S, N\}$, is

$$\begin{aligned}e_2^k(b_1) &= \beta \left(\int_{b_1}^{\bar{b}} b dF(b) - b_1(1 - F(b_1)) \right) \\ &= \beta \int_{b_1}^{\bar{b}} (b - b_1) dF(b).\end{aligned}$$

In the remainder of this proof we denote the censored expected difference between a new offer and an old offer conditional on the old offer granting payoff b_1 with

$C(b_1) = \int_{b_1}^{\bar{b}} (b - b_1) dF(b)$. Note that $C(0) = \mathbb{E}(b)$, while $C(b_1) < \mathbb{E}(b) \forall b_1 > 0$. The agent only accepts an offer in period 2 if the following condition is satisfied:

$$w_2^k - \frac{1}{2}e_2^k(b_1)^2 + \beta \left[e_2^k(b_1) \left(\int_{b_1}^{\bar{b}} b dF(b) + b_1 F(b_1) \right) + (1 - e_2^k(b_1))b_1 \right] \geq b_1 + \beta b_1$$

$$\Rightarrow w_2^k(b_1) = b_1 + \frac{1}{2}e_2^k(b_1)^2 - \beta e_2^k(b_1)C(b_1) = b_1 - \frac{1}{2}[\beta C(b_1)]^2.$$

For the period-2 wage and effort of the time-consistent agent, it is sufficient to replace β with 1 in the above expressions, hence $e_2^{TC}(b_1) = C(b_1)$ and $w_2^{TC}(b_1) = b_1 - \frac{1}{2}C(b_1)^2$.

Taking those values as given, it is instructive to calculate the long-run realized period-2 utility for a job offer b_1 . Note that, for the time-consistent agent, this realized long-run utility is necessarily equal to the outside option, i.e. $2b_1$. Since the equilibrium outcomes for the naive and the sophisticated agents are the same, so is the long-run period-2 utility, namely:

$$2b_1 + \beta(1 - \beta)C(b_1)^2 > 2b_1.$$

For the analysis of the equilibrium contract in period 1, we start with the **naive agent**. Since the naive agent believes to act like a time-consistent agent, i.e. $\tilde{e}_2^N(b_1) = e_2^{TC}(b_1)$ and $\tilde{w}_2^N(b_1) = w_2^{TC}(b_1)$, he also believes that his utility from employment in period 2 will be that of a time consistent agent. Hence his perceived utility from employment in the first period is given by:

$$U_1^N = w_1 - \frac{1}{2}e_1^2 + \beta [e_1 \mathbb{E}(2b) + (1 - e_2)0] \quad (3)$$

$$= w_1 - \frac{1}{2}e_1^2 + e_1 2\beta \mathbb{E}(b) \quad (4)$$

Consequently, optimal search and wage for the naive agent will be:

$$e_1^N = 2\beta \mathbb{E}(b) \quad \text{and} \quad w_1^N = \frac{1}{2}(e_1^N)^2 - \beta e_1^N 2\mathbb{E}(b) = -\frac{1}{2}(2\beta \mathbb{E}(b))^2.$$

This leads to the following long-run realized utility from employment:

$$\hat{U}_0^N = w_1^N - \frac{1}{2}(e_1^N)^2 + [e_1^N (\mathbb{E}(2b) + \beta(1 - \beta)\mathbb{E}(C(b)^2)) + (1 - e_1^N)\beta(1 - \beta)\mathbb{E}(b)^2]$$

$$= (1 - \beta)e_1^N 2\mathbb{E}(b) + e_1^N \beta(1 - \beta)\mathbb{E}(C(b)^2) + (1 - e_1^N)\beta(1 - \beta)\mathbb{E}(b)^2 > 0$$

In contrast, the **sophisticated agent** correctly anticipates his future present

bias, hence also that his long-run utility will be larger than the outside option:

$$U_1^S = w_1 - \frac{1}{2}e_1^2 + \beta [e_1 (\mathbb{E}(2b) + \beta(1 - \beta)\mathbb{E}(C(b)^2)) + (1 - e_2)\beta(1 - \beta)\mathbb{E}(b)^2] \quad (5)$$

Therefore, optimal effort of the sophisticated agent in period 1 is:

$$\begin{aligned} e_1^S &= \beta [\mathbb{E}(2b) + \beta(1 - \beta)\mathbb{E}(C(b)^2) - \beta(1 - \beta)\mathbb{E}(b)^2] \\ &= 2\beta\mathbb{E}(b) - \beta(1 - \beta) [\mathbb{E}(b)^2 - \mathbb{E}(C(b)^2)] < 2\beta\mathbb{E}(b) = e_1^N \end{aligned}$$

Moreover, the principal offers him the following wage:

$$\begin{aligned} w_1^S &= \frac{1}{2}(e_1^S)^2 - \beta e_1^S 2\mathbb{E}(b) - \beta (e_1^S \beta(1 - \beta)\mathbb{E}(C(b)^2) \\ &\quad + (1 - e_1^S)\beta(1 - \beta)\mathbb{E}(b)^2) \\ &< \frac{1}{2}(e_1^S)^2 - \beta e_1^S 2\mathbb{E}(b) \end{aligned}$$

Summarizing, the principal can push the period-1 wage of the sophisticated agent below the (negative) net surplus of search not only due to the “extra” utility of assessing period-2 search from the perspective of period 1 for the case that search has not been successful, but also for the case of receiving a better offer. Thus, the long-run realized utility of the sophisticated agent is:

$$\begin{aligned} \hat{U}_0^S &= w_1^S - \frac{1}{2}(e_1^S)^2 + e_1^S [\mathbb{E}(2b) + \beta(1 - \beta)\mathbb{E}(C(b)^2)] + (1 - e_1^S)\beta(1 - \beta)\mathbb{E}(b)^2 \\ &= (1 - \beta)e_1^S 2\mathbb{E}(b) + (1 - \beta)e_1^S \beta(1 - \beta)\mathbb{E}(C(b)^2) + (1 - \beta)(1 - e_1^S)\beta(1 - \beta)\mathbb{E}(b)^2 \\ &< (1 - \beta)e_1^S 2\mathbb{E}(b) + e_1^S \beta(1 - \beta)\mathbb{E}(C(b)^2) + (1 - e_1^S)\beta(1 - \beta)\mathbb{E}(b)^2 \\ &< (1 - \beta)e_1^N 2\mathbb{E}(b) + e_1^N \beta(1 - \beta)\mathbb{E}(C(b)^2) + (1 - e_1^N)\beta(1 - \beta)\mathbb{E}(b)^2 = \hat{U}_0^N \end{aligned}$$

The inequality in the third line follows from $(1 - \beta) < 1$ and the inequality in the second line follows from $e_1^S < e_1^N$ and:

$$2\mathbb{E}(b) + \beta\mathbb{E}(C(b)^2) > \beta\mathbb{E}(b)^2$$

as $\mathbb{E}(b) < 1$. Note that from the second line it is evident that $\hat{U}_0^S > 0$.

Finally, for the time-consistent agent it follows that $e_1^{TC} = 2\mathbb{E}(b) > e_1^N$ and $\hat{U}_0^{TC} = U_1^{TC} = 0$, which establishes the proposition. \blacksquare

A.6.2 Proof of Proposition 4 (b as Performance-Based Bonus)

Plugging $e_1^S = e_1^N = \beta b$ and $b_1^S = b_1^N = \theta/(2 - \beta)$ into an agent's first-period utility and setting it equal to zero yields

$$w_1^S = -\frac{\left(\beta \frac{\theta}{2-\beta}\right)^2}{2} - \beta^2 \left(\frac{\theta}{2-\beta}\right)^2 (1-\beta) = -\left(\beta \frac{\theta}{2-\beta}\right)^2 \left[\frac{3-2\beta}{2}\right]$$

and $w_1^N = w_2^N (= -\frac{1}{2} \left(\frac{\beta}{2-\beta}\theta\right)^2)$.

Therefore,

$$\begin{aligned} \hat{U}_0^S &= w_1^S + e_1^S b_1^S - c(e_1^S) + w_2^S + e_1^S b_2^S - c(e_1^S) \\ &= \frac{(1-\beta)\beta\theta^2}{(2-\beta)} \\ \hat{U}_0^N &= 2\frac{(1-\beta)\beta\theta^2}{(2-\beta)^2} > \hat{U}_0^S \end{aligned}$$

Profits are

$$\begin{aligned} \Pi^N &= \theta^2 \frac{(3-2\beta)\beta}{(2-\beta)^2} \\ \Pi^S &= \frac{\theta^2\beta(3-2\beta)}{(2-\beta)^2} + \beta^2 \left(\frac{\theta}{2-\beta}\right)^2 (1-\beta) > \Pi^N \\ \Pi^{TC} &= \theta^2, \end{aligned}$$

with,

$$\begin{aligned} \Pi^{TC} &> \Pi^S \\ \Leftrightarrow 4 - 2\beta - \beta(1+\beta) &> 0. \end{aligned}$$

■

A.7 Extensions

A.7.1 Proof of Proposition 5 (Employment Protection)

The arguments in the text deliver that

$$\begin{aligned}\tilde{w}_2^N &= \min \left\{ -\frac{1}{2}(\beta b)^2, k - \frac{1}{2}b^2 \right\} \\ w_2^N &= -\frac{1}{2}(\beta b)^2.\end{aligned}$$

The naive agent's perceived period-1 utility equals

$$\begin{aligned}\tilde{U}_1^N &= w_1^N - \frac{1}{2}(e_1^N)^2 + \beta \left[e_1^N b + \tilde{w}_2^N + \frac{1}{2}b^2 \right] \\ &= w_1^N + \frac{1}{2}(\beta b)^2 + \beta \left(\tilde{w}_2^N + \frac{1}{2}b^2 \right),\end{aligned}$$

thus

$$w_1^N = \begin{cases} -\frac{(b\beta)^2}{2} - \beta k & \text{if } k < \frac{1}{2}b^2(1 - \beta^2) \\ -\frac{b^2\beta(1 + \beta(1 - \beta))}{2} & \text{if } k \geq \frac{1}{2}b^2(1 - \beta^2). \end{cases}$$

Clearly, if $w_1^N = -b^2\beta(1 + \beta(1 - \beta))/2$, the naive agent's first period wage is lower than the sophisticated agent's first-period wage, $w_1^S = -(\beta b)^2(3 - 2\beta)/2$. However, it is still larger than the wage of the time-consistent agent $w_1^{TC} = -\frac{1}{2}b^2$.

Moreover, since $w_2^N = -(\beta b)^2/2$ and $\hat{U}_0^N = w_1^N + w_2^N + \beta b^2(2 - \beta)$,

$$\hat{U}_0^N = \begin{cases} -\beta k + 2\beta b^2(1 - \beta) & \text{if } k < \frac{1}{2}b^2(1 - \beta^2) \\ \frac{b^2\beta((2 - \beta)^2 - 1)}{2} & \text{if } k \geq \frac{1}{2}b^2(1 - \beta^2) \end{cases}$$

It immediately follows that, if $\hat{U}_0^N = b^2\beta((2 - \beta)^2 - 1)/2$, it is lower than the sophisticated agent's long-run utility, $\hat{U}_0^S = (2 - \beta)\beta b^2(1 - \beta)$, but larger than the level of the time-consistent agent. ■

A.7.2 Proof of Proposition 6 (Minimum Wage)

To begin with, note that the principal's lack of commitment rules out the use of a firing threat to provide incentives, which might otherwise be optimal to reduce the agent's first-period rent (for the same reasons as in moral hazard problems with limited liability).

Now, recall that, without a minimum wage,

$$\begin{aligned}\tilde{w}_2^N &= -\frac{1}{2}b^2 \\ < w_1^N = w_2^N = w_2^S &= -\frac{1}{2}(\beta b)^2.\end{aligned}$$

Assume $\bar{w} \geq -\frac{1}{2}b^2$, hence $\tilde{w}_2^N = \bar{w}$.

Furthermore, recall that the naive agent's perceived period-1 utility equals

$$\begin{aligned}\tilde{U}_1^N &= w_1^N - \frac{1}{2}(e_1^N)^2 + \beta \left[e_1^N b + \left(\bar{w} + \frac{1}{2}b^2 \right) \right] \\ &= w_1^N + \frac{1}{2}(\beta b)^2 + \beta \left(\bar{w} + \frac{1}{2}b^2 \right),\end{aligned}$$

thus w_1^N is the lowest feasible wage that satisfies $\tilde{U}_1^N \geq 0$, or

$$w_1 = \max \left\{ \bar{w}; - \left[\frac{1}{2}(\beta b)^2 + \beta \left(\bar{w} + \frac{1}{2}b^2 \right) \right] \right\}.$$

Moreover,

$$\begin{aligned}\bar{w} &< - \left[\frac{1}{2}(\beta b)^2 + \beta \left(\bar{w} + \frac{1}{2}b^2 \right) \right] \\ \Leftrightarrow \bar{w} &< \bar{w}^* \equiv -\frac{1}{2}\beta b^2.\end{aligned}$$

Assume this condition holds (note that $\bar{w}^* < -\frac{1}{2}(\beta b)^2$, i.e., $w_2^N > \bar{w}^*$), then

$$\begin{aligned}\hat{U}_0^N &= w_1^N - \frac{1}{2}(e_1^N)^2 + e_1^N b + w_2^N - \frac{1}{2}(e_2^N)^2 + e_2^N b \\ &= w_1^N + \beta b^2 \frac{(4-3\beta)}{2} \\ &= \frac{(3-4\beta)}{2} \beta b^2 - \beta \bar{w},\end{aligned}$$

which is decreasing in \bar{w} .

Now, assume $\bar{w} \geq \bar{w}^*$, hence $w_1 = \bar{w}$.

For $\bar{w} \in [\bar{w}^*, -\frac{1}{2}(\beta b)^2]$,

$$\hat{U}_0^N = \bar{w} + \beta b^2 \frac{(4-3\beta)}{2},$$

for $\bar{w} > -\frac{1}{2}(\beta b)^2$,

$$\hat{U}_0^N = 2\bar{w} + \beta b^2 (2-\beta),$$

which both are increasing in \bar{w} .

The results for the time-inconsistent but sophisticated and time-consistent agents follow: They are only affected by a binding minimum wage; if the minimum wage exceeds w_1^S (w_1^{TC}), the respective first-period wages without a minimum wage, U_1^S

(U_1^{TC}) and consequently \hat{U}_0^S (\hat{U}_1^{TC}) go up. ■

A.7.3 Asymmetric Information

Proof of Proposition 7a Recall the second-period wage and first- as well as second-period search effort of a time-inconsistent agent:

$$w_2^k = -\frac{1}{2}(\beta b)^2 < 0 \text{ and } e_2^k = e_1^k = \beta b,$$

where $k \in \{S, N\}$. Since the sophisticated agent is fully aware of the future contract terms, his reservation wage will not change compared to the symmetric information case. The same is true for the naive agent: Since he believes everyone, including himself, to be time-consistent in the future, he anticipates a wage $\tilde{w}_2^N = -\frac{1}{2}b^2$ and hence also has the same reservation wage compared to the symmetric information case, w_1^N .

From the symmetric information case we know that $w_1^S < w_1^N$. Thus, the principal faces two options: Either he offers w_1^N and both types of agents accept, or he offers w_1^S and only the sophisticated agent accepts. Employing both types of agents in the first period is optimal if:

$$3\pi - w_1^N - w_2^k - w_3^k \geq \alpha(3\underline{\pi}) + (1 - \alpha)(3\pi - w_1^S - w_2^k - w_3^k)$$

Rearranging the inequality yields:

$$\begin{aligned} \alpha(3(\pi - \underline{\pi}) - w_1^S - w_2^k) &\geq w_1^N - w_1^S \\ \alpha &\geq \frac{w_1^N - w_1^S}{3(\pi - \underline{\pi}) - w_1^S - w_2^k} = \bar{\alpha} \end{aligned}$$

Clearly, $\bar{\alpha}$ is strictly larger than 0 since $w_1^S < w_1^N$. Moreover, it is strictly smaller than 1 since $w_1^N < 0$, $w_2^k < 0$ and $\pi > \underline{\pi}$.

Now, either way the principal's expected profits are lower than under symmetric information: If $\alpha \geq \bar{\alpha}$, then she employs both agents at a wage w_1^N and cannot extract the sophisticated agent's rent (but there is always a small share of sophisticated agents because $\alpha < 1$). If $\alpha < \bar{\alpha}$, she decides to extract the sophisticated agent's rent, but at the cost of missing the employment opportunity when she is matched with a naive agent. ■

Proof of Proposition 7b Remember the second-period wage and first- as well as second-period search effort of a time-inconsistent agent:

$$w_2^k = -\frac{1}{2}(\beta b)^2 < 0 \text{ and } e_2^k = e_1^k = \beta b$$

where $k \in \{S, N\}$. Since the sophisticated agent is fully aware of the future contract terms, his reservation wage will not change compared to the symmetric information case. Now, however, the naive agent perceives all other agents except himself to be time-inconsistent in the future, thus he anticipates a wage $w_2^S = -\frac{1}{2}(\beta b)^2$. Despite expecting the same wage as the sophisticated agent, the naive agent's reservation wage in period 1 is lower because he anticipates a larger effort $\tilde{e}_2^N = b$:

$$\begin{aligned} U_1^N &= w_1 - \frac{1}{2}(\beta b)^2 + \beta \left[\beta b^2 - \frac{1}{2}(\beta b)^2 - \frac{1}{2}b^2 + b^2 \right] \geq 0 \\ \Rightarrow \hat{w}_1^N &= -\frac{1}{2}(\beta b)^2 - \frac{1}{2}\beta(1 - \beta^2)b^2 < -\frac{1}{2}(\beta b)^2 - \beta^2(1 - \beta)b^2 = w_1^S \end{aligned}$$

Hence, the principal faces again a trade-off: Either he offers \hat{w}_1^N and only the naive agent accepts, or he offers w_1^S and both types of agents accept. Employing only naive agents is optimal if:

$$\alpha(3\pi - w_1^N - w_2^k - w_3^k) + (1 - \alpha)3\underline{\pi} > 3\pi - w_1^S - w_2^k - w_3^k$$

Rearranging the inequality yields:

$$\alpha > \frac{3(\pi - \underline{\pi}) - w_1^S - w_2^k}{3(\pi - \underline{\pi}) - w_1^N - w_2^k} = \bar{\alpha}$$

Clearly, $\bar{\alpha}$ is strictly larger than 0. Moreover, it is strictly smaller than one since $w_1^N < w_1^S < 0$.

The principal's profits are larger than under symmetric information in both cases. First, consider the case when $\alpha \leq \bar{\alpha}$. Then the principal can employ both agents at the reservation wage of the sophisticated agent $w_1^S > w_1^N$, where w_1^N is the reservation wage of the naive agent under symmetric information. Clearly, she is better off than under symmetric information in expectation because she can extract the same rent from both, sophisticated and naive agent instead of only the sophisticated (and there is always a small share of naive agents because $\alpha > 0$). Now, when $\alpha > \bar{\alpha}$, the principal decides to exclude the sophisticated agent because her expected profits increase *above* what would be possible by employing both type of agents at w_1^S . ■

A.8 Robustness

A.8.1 Proof of Lemma 4 (Partial Naiveté)

A partially naive agent perceives his first-period utility to be (already taking into account \tilde{w}_2 and \tilde{e}_2)

$$\tilde{U}_1^{PN} = w_1 - \frac{1}{2}(e_1)^2 + \beta \left[e_1 b + \tilde{\beta}(1 - \tilde{\beta})b^2 \right].$$

Since, as before, the optimal effort level is given by $e_1^{PN} = \beta b$, the reserve wage is given by

$$\begin{aligned} w_1^{PN} &= \frac{1}{2}(e_1^{PN})^2 - \beta \left[e_1^{PN} b + \tilde{\beta}(1 - \tilde{\beta})b^2 \right] \\ &= -\frac{1}{2}(\beta b)^2 - \beta \tilde{\beta}(1 - \tilde{\beta})b^2 < -\frac{1}{2}(\beta b)^2 = w_2^{PN} \end{aligned}$$

and

$$\frac{dw_1^{PN}}{d\tilde{\beta}} = -\beta(1 - 2\tilde{\beta}) > 0 \text{ since } \tilde{\beta} > \beta \geq \frac{1}{2}$$

■

A.8.2 Proof of Proposition 7 (Discounting between each Step)

We first compute outcomes for the *naive* agent. The discussion in the text implies that $e_2^N = \beta b$, $\tilde{e}_2^N = b$, thus he is willing to accept a wage $w_2^N = -\beta b^2/2$, whereas (in period 1) he anticipates a wage $\tilde{w}_2^N = -b^2/2$. Equivalently, $e_1^N = \beta b$ and $\tilde{e}_1^N = b$.

His first-period wage is determined by

$$w_1^N + \beta \left(\tilde{e}_1^N b - \frac{(\tilde{e}_1^N)^2}{2} + \tilde{w}_2^N + \tilde{e}_2^N b - \frac{(\tilde{e}_2^N)^2}{2} \right) = 0,$$

therefore

$$w_1^N = -\beta \frac{b^2}{2},$$

and

$$\begin{aligned} \hat{U}_0^N &= w_1^N + e_1^N b - \frac{(e_1^N)^2}{2} + w_2^N + e_2^N b - \frac{(e_2^N)^2}{2} \\ &= \beta b^2 (1 - \beta) > 0. \end{aligned}$$

Next, we compute outcomes for the *sophisticated* agent. His effort levels are

$e_2^S = e_1^S = \beta b$, his second-period wage solves $w_2^S + \beta (-(e_2^S)^2/2 + e_2^S b) = 0$, thus

$$w_2^S = -\frac{2-\beta}{2}(\beta b)^2 < 0.$$

Note that $w_2^S > w_2^N = -\beta b^2/2$.

His first-period wage is given by $w_1 + \beta (-(e_1^S)^2/2 + e_1^S b - \frac{2-\beta}{2}(\beta b)^2 - (e_2^S)^2/2 + e_2^S b) = 0$, thus

$$w_1^S = -\frac{(2-\beta)^2}{2}(\beta b)^2 < w_2^S.$$

Therefore, his long-run utility equals

$$\begin{aligned} \hat{U}_0^S &= w_1^S + e_1^S b - \frac{(e_1^S)^2}{2} + w_2^S + e_2^S b - \frac{(e_2^S)^2}{2} \\ &= \beta b^2 \frac{(2-\beta)^2(1-\beta)}{2} > 0. \end{aligned}$$

It follows that

$$\begin{aligned} \hat{U}_0^N &> \hat{U}_0^S \\ \Leftrightarrow 2 &> (2-\beta)^2 \end{aligned}$$

which holds if and only if β is sufficiently large. ■

B Supplementary Appendix

B.1 Endogenous Bonus and Employment Protection

Introducing (limited) commitment has further implications for our application of an endogenous bonus because the principal can also use her commitment to already set second-period incentives in the first period. Here, we first focus on the sophisticated agent. For him, the principal might want to commit to a higher second-period bonus because higher effort increases the “extra rent” of second-period effort from the perspective of period 1. To see that, let us first assume that the principal can fully commit to second-period outcomes, and in particular the bonus b_2^S . Then, it would still be optimal to set w_2^S to fully extract the second-period rent, i.e., $w_2^S = -(\beta b_2^S)^2/2$. Moreover, first-period bonus and effort would be the same as without commitment, thus $U_1^S = w_1^S + \frac{1}{2} \left(\frac{\beta \theta}{(2-\beta)} \right)^2 + (\beta b_2^S)^2 (1-\beta)$, which delivers

$$w_1^S = -\frac{1}{2} \left(\frac{\beta\theta}{(2-\beta)} \right)^2 - (\beta b_2^S)^2 (1-\beta).$$

Therefore, the principal maximizes

$$\Pi_1 = \frac{\beta\theta^2}{2(2-\beta)} + (\beta b_2^S)^2 \frac{(3-2\beta)}{2} + \beta b_2^S (\theta - b_2^S),$$

and the optimal b_2^S equals

$$b_2^S = \frac{\theta}{(2-\beta(3-2\beta))}.$$

This is larger than $\theta/(2-\beta)$, the second-period bonus without commitment.

All this yields

Proposition 8 *Assume that, with an endogenous bonus, the principal can commit to future contracts but deviate at cost k . With a sophisticated agent, she offers a second-period bonus which exceeds the optimal myopic level. Moreover, the principal's long-run profits and the agent's long-run utility (weakly) increase in k*

Proof:

We have shown that the principal would like to commit to a second-period bonus which exceeds the myopically optimal level, $\theta/(2-\beta)$. Generally, a bonus b_2^S can only be credibly promised if

$$\beta b_2^S (\theta - b_2^S) + \frac{(\beta b_2^S)^2}{2} \geq \frac{\beta\theta^2}{2(2-\beta)} - k,$$

where $\pi_2 = \beta b_2^S (\theta - b_2^S) + (\beta b_2^S)^2 / 2$ are second-period profits for a general bonus, and $\beta\theta^2 / [2(2-\beta)]$ is the maximum level of second-period profits.

If the bonus $b_2^S = \theta / (2 - \beta(3 - 2\beta))$ (which is the optimal long-term bonus, as derived in the main text) satisfies this condition, it is offered. Otherwise, the bonus is determined by the binding condition, with

$$\frac{db_2^S}{dk} = -\frac{1}{\beta(\theta - 2b_2^S) + \beta^2 b_2^S} > 0,$$

where the inequality follows from the denominator being negative for $b_2^S > \theta/(2-\beta)$.

It also follows that long-term profits increase in k as long as $b_2^S < \theta / (2 - \beta(3 - 2\beta))$.

To determine the agent's long-run utility, we plug

$$w_1^S = -\frac{1}{2} \left(\frac{\beta\theta}{(2-\beta)} \right)^2 - (\beta b_2^S)^2 (1-\beta)$$

into

$$\begin{aligned} \tilde{U}_0^S &= w_1^S + \frac{\beta\theta^2}{2(2-\beta)} + \beta (b_2^S)^2 (1-\beta) \\ &= \frac{\beta\theta^2 (1-\beta)}{(2-\beta)^2} + \beta (b_2^S)^2 (1-\beta)^2. \end{aligned}$$

This increases in b_2^S and consequently in k (as long as $b_2^S < \theta / (2 - \beta(3 - 2\beta))$). ■

A higher “extra rent” of second-period effort from the perspective of the first period benefits the principal who can extract it. It also benefits the agent, though, because the principal's extraction is only complete from the perspective of the first, not from the perspective of earlier periods (where there is no discounting between periods 1 and 2).

Now, we move on and analyze the optimal long-term contract for the naive agent if the principal has limited commitment. Then, as before she would like to commit to a higher second-period wage and consequently reduce w_1^N . She will not offer the perceived surplus-maximizing bonus $b_2 = \theta$, though, and instead *not* specify a second-period bonus at all (or offer the actually optimal level $b_2^N = \theta / (2 - \beta)$). This is because the agent anticipates a second-period bonus $\tilde{b}_2^N = \theta$ *anyway*, thus expects any lower bonus to be renegotiated upwards (to which she would agree because a higher bonus increases the agent's rent). If, however, the principal specified $\tilde{b}_2^N = \theta$ already in the first-period contract, she would have to compensate the agent (or pay k) if she wanted to reduce b_2^N to the actually optimal level $\theta / (2 - \beta)$. Thus, the size of k determines outcomes as specified by the following proposition.

Proposition 9 *Assume that, with an endogenous bonus, the principal can commit to future contracts but deviate at cost k . With a naive agent, she offers a second-period wage that exceeds the profit-maximizing level as perceived by the agent's first-period self. The principal's profits (weakly) increase and the agent's long-run utility (weakly) decrease in k . If k is sufficiently large, the agent's long-run utility is zero.*

Proof:

First-period bonus and effort levels are as without commitment. Moreover, let us assume that, in the first period, the principal does not promise a second-period bonus \tilde{b}_2^N (which is weakly optimal), only a wage \tilde{w}_2^N . Then, the agent expects $\tilde{b}_2^N = \theta$ to be offered at the beginning of the second period (and anticipates an effort choice $\tilde{e}_2^N = \theta$). Thus, for a given (credible) \tilde{w}_2^N , the agent's perceived first-period utility equals

$$\tilde{U}_1^N = w_1^N + \frac{1}{2} \left(\frac{\beta\theta}{(2-\beta)} \right)^2 + \beta \left(\tilde{w}_2^N + \frac{\theta^2}{2} \right), \text{ thus}$$

$$w_1^N = -\frac{1}{2} \left(\frac{\beta\theta}{(2-\beta)} \right)^2 - \beta \left(\tilde{w}_2^N + \frac{\theta^2}{2} \right),$$

and the principal's profits are

$$\begin{aligned} \Pi_1 &= e_1^N (\theta - b_1^N) - w_1^N + e_2^N (\theta - b_2^N) - w_2^N \\ &= \beta\theta^2 \frac{8 - 7\beta + \beta^2}{2(2-\beta)^2} + \beta\tilde{w}_2^N - w_2^N, \end{aligned}$$

where we have already taken into account that $b_2^N = \theta/(2-\beta)$ and $e_2^N = \beta b_2^N = \beta\theta/(2-\beta)$. As in the case with exogenous bonus, the principal would like to commit to the wage she will eventually offer and that just extracts the second period surplus, i.e., to $\tilde{w}_2 = w_2^N = -\frac{(\beta\theta)^2}{2(2-\beta)^2}$. However, she is restricted by the condition $\tilde{w}_2^N \leq k - \frac{\theta^2}{2}$, where $-\theta^2/2$ is the wage the agent would expect without any commitment. All this implies

$$\begin{aligned} \tilde{w}_2^N &= \min \left\{ -\frac{(\beta\theta)^2}{2(2-\beta)^2}, k - \frac{\theta^2}{2} \right\} \\ w_2^N &= -\frac{(\beta\theta)^2}{2(2-\beta)^2}. \end{aligned}$$

Thus, if $k \geq \frac{2(1-\beta)}{(2-\beta)^2}\theta^2$ and $\tilde{w}_2^N = w_2^N = -\frac{(\beta\theta)^2}{2(2-\beta)^2}$,

$$\Pi_1 = \beta\theta^2 \frac{8 - 7\beta + \beta^2}{2(2-\beta)^2} + \frac{(\beta\theta)^2}{2(2-\beta)^2} (1 - \beta)$$

and

$$\begin{aligned} \hat{U}_0^N &= \frac{\beta^2\theta^2}{2} \frac{(1-\beta)}{(2-\beta)^2} - \beta\tilde{w}_2^N + w_2^N \\ &= 0. \end{aligned}$$

Otherwise, $\tilde{w}_2^N = k - \frac{\theta^2}{2}$, and

$$\begin{aligned}\Pi_1 &= \beta\theta^2 \frac{8 - 7\beta + \beta^2}{2(2 - \beta)^2} + \beta\tilde{w}_2^N - w_2^N \\ &= \beta\theta^2 \frac{8 - 6\beta + \beta^2}{2(2 - \beta)^2} - \beta\frac{\theta^2}{2} + \beta k,\end{aligned}$$

as well as

$$\begin{aligned}\hat{U}_0^N &= \frac{\beta^2\theta^2}{2} \frac{(1 - \beta)}{(2 - \beta)^2} - \beta\tilde{w}_2^N + w_2^N \\ &= \beta\theta^2 \frac{2(1 - \beta)}{(2 - \beta)^2} - \beta k.\end{aligned}$$

■

As with an exogenous bonus, the agent is harmed by a higher termination cost k because those allow the principal to credibly promise a higher second-period wage and consequently to reduce w_1^N . In addition, the agent overestimates second-period incentives and his second-period effort which, for a given expected second-period wage \tilde{w}_2^N , allows the principal to reduce w_1^N even further. If k is sufficiently high such that the principal can already commit to the actually paid second-period wage, the wrong anticipation of a high second-period rent lets the agent's long-run utility completely vanish.

B.2 Endogenous Bonus and Infinite Time Horizon

Before deriving an equilibrium in which the principal endogenously commits to a bonus exceeding the myopic optimum, note that it can be shown that, for any number of periods in a *finite* time horizon, the principal would set the bonus $b = \theta/(2 - \beta)$ in every period, resulting in effort $e = \beta\delta\theta/(2 - \beta)$. These are the same values as in the main part, incorporating the additional exponential discounting. This holds for the naive agent who does not anticipate any future rent anyway, but also for the sophisticated agent. There, any promise by the principal to pay a higher bonus in future periods would not be credible because of a standard backwards induction argument: In the last period, there is a unique equilibrium with $b = \theta/(2 - \beta)$, thus also in the second to last period, and so on.

With an infinite time horizon, an equilibrium with $b = \theta/(2 - \beta)$ naturally exists as well. However, we now show that there exists another equilibrium in which the principal can increase her long-run utility by promising a higher bonus in all periods $t \geq 2$. Such an equilibrium could be sustained by the following strategy:

The principal offers $b > \theta/(2 - \beta)$ in every period $t \geq 2$, together with a wage that completely extracts the agent's rent from today's perspective. If the principal deviates in any period t , she offers the spot bonus $\theta/(2 - \beta)$ forever thereafter, which is incorporated by today's wage.

Proposition 10 *With an infinite time horizon and an endogenous bonus, an equilibrium with the following characteristics exists:*

- *For the sophisticated agent, the principal offers the myopic bonus $b_1^S = \theta/(2 - \beta)$ in the first period, and a bonus $b^S > b_1^S$ in all subsequent periods. The agent's long-run utility is higher than if the myopic bonus is paid in every period.*
- *For the naive agent, the principal offers the myopic bonus $b^N = \theta/(2 - \beta)$ in every period. The naive agent's long-run utility can be larger or smaller than the sophisticated agent's long-run utility.*

Proof:

We start with the sophisticated agent and first derive a bonus b_1^S for the first and a bonus b^S for all future periods (together with optimal wages w_1^S and w^S) which maximize the principal's first-period profit stream. Then, we show that promising these profit-maximizing bonuses is credible if and only if δ is sufficiently large.

In any future period, effort equals $e^S = \beta\delta b^S$, thus the wage w^S solves

$$w^S - \frac{(e^S)^2}{2} + \beta\delta \left(e^S b^S + \frac{w^S - \frac{(e^S)^2}{2} + \delta e^S b^S}{1 - \delta} \right) = 0$$

$$\Rightarrow w^S = -\frac{(\beta\delta b^S)^2}{2} \left(\frac{1 + \delta(1 - \beta)}{1 - \delta(1 - \beta)} \right)$$

The agent's first period utility equals

$$U_1^S = w_1^S - \frac{(e_1^S)^2}{2} + \beta\delta \left(e_1^S b_1^S + \frac{w^S - \frac{(e^S)^2}{2} + \delta e^S b^S}{1 - \delta} \right)$$

$$= w_1^S + \frac{(\beta\delta b_1^S)^2}{2} + \delta \frac{(\beta\delta b^S)^2 (1 - \beta)}{(1 - \delta(1 - \beta))} = 0$$

For later use, note that the principal's long-run utility is

$$\begin{aligned}\hat{U}^S &= w_1^S - \frac{(e_1^S)^2}{2} + \delta \left(e_1^S b_1^S + \frac{w^S - \frac{(e^S)^2}{2} + \delta e^S b^S}{1 - \delta} \right) \\ &= \beta (\delta b_1^S)^2 (1 - \beta) + (\delta b^S)^2 \frac{\delta \beta (1 - \beta)^2}{(1 - \delta (1 - \beta))},\end{aligned}$$

which is strictly positive and increasing in b^S (as well as in b_1^S).

In the next step, we compute the levels of b_1^S and b^S that maximize the principal's profits,

$$\begin{aligned}\Pi_1 &= -w_1^S + e_1^S \delta (\theta - b_1^S) + \delta \frac{e \delta (\theta - b^S) - w^S}{1 - \delta} \\ &= \beta \delta^2 b_1^S \left(\theta - b_1^S \frac{2 - \beta}{2} \right) + \delta \frac{\beta \delta^2 b^S (\theta - b^S) + \frac{(\beta \delta b^S)^2 [2(1 - \beta) + 1 - \delta(1 - \beta)]}{2(1 - \delta(1 - \beta))}}{1 - \delta}.\end{aligned}$$

First-order conditions yield

$$\begin{aligned}b_1^S &= \frac{\theta}{(2 - \beta)} \\ b^S &= \frac{\theta (1 - \delta (1 - \beta))}{(2 - \beta) (1 - \delta (1 - \beta)) - 2\beta (1 - \beta)},\end{aligned}$$

with $b^S > b_1^S$.

Promising b^S is credible if a deviation to the myopic optimal $\tilde{b} = \theta / (2 - \beta)$ is not optimal. Taking $e = \beta \delta b$ into account, this holds if

$$\frac{\beta \delta^2 b^S (\theta - b^S) - w^S}{1 - \delta} \geq \frac{\beta \delta^2 \tilde{b} (\theta - \tilde{b}) - \tilde{w}}{1 - \delta}, \quad (6)$$

where \tilde{w} is set to keep the agent's utility at zero.

It can be shown that condition (6) holds for δ sufficiently close to 1, and is violated if δ is small. If it is violated, b^S is determined by the binding constraint (6). In any case, $b^S \geq \frac{\theta(1 - \delta(1 - \beta))}{2(1 - \delta(1 - \beta)) - \beta(1 + \delta(1 - \beta))}$, which is the bonus that maximizes $\beta \delta^2 b (\theta - b) - w$, and thus exceeds b_1^S .

For the naive agent, a contract that differs from the myopic optimum would only be possible if a deviation generated lower continuation profits. Since the principal cannot do worse than the myopic optimum, her continuation profits after a deviation, assessed from the perspective of the naive agent, would involve the optimal contract for the time-consistent agent, with a bonus $\tilde{b}^N = \theta$ and a wage

$\tilde{w}^N = -(\delta\theta)^2/2$. No other contract could generate higher profits and would be accepted by the naive agent's perceived future self, thus the optimal continuation contract (from the perspective of the naive agent) after every history would be the myopic optimum. But this implies that no other contract than the myopic optimum can be offered in any period.

Finally, plugging optimal bonuses into an agent's long-run utility yields

$$\begin{aligned}\hat{U}^N &= \frac{\beta \left(\frac{\delta\theta}{(2-\beta)}\right)^2 (1-\beta)}{1-\delta} \\ \hat{U}^S &= \beta \left(\frac{\delta\theta}{(2-\beta)}\right)^2 (1-\beta) + (\delta b^S)^2 \frac{\delta\beta(1-\beta)^2}{(1-\delta(1-\beta))} \\ &\Rightarrow \hat{U}^N \geq \hat{U}^S \\ &\Leftrightarrow 4\beta(1-\beta)^2 \\ &\geq (2-\beta)(1-\delta(1-\beta))(2-3\beta).\end{aligned}$$

This holds for β sufficiently large (then, the right hand side is negative), and is violated if β is small. ■

With the sophisticated agent, the principal can use the infinite time horizon to credibly promise a bonus which exceeds the myopic optimum. This increases the agent's "extra rent" from his future time inconsistency from which both, principal and agent, benefit in the long run. The promise is credible (in every period, the principal would actually like to offer b_1^S in the present and b^S in the future) if a deviation is not too tempting. In the proof to Proposition (10), we show that a $b^S > \theta/(2-\beta)$ can always be implemented. Whether this also holds for the bonus maximizing the principal's long-run profits depends on the discount factor δ . Only if it is high enough, a permanent reversion to the myopic optimum in the future is sufficiently unattractive for a deviation today being deterred.